

VISIT...

LANZAROTE
Caliente.COM

KIRK-OTHMER

ENCYCLOPEDIA OF

CHEMICAL

TECHNOLOGY

EDITING - by **WATCHER**

VOLUME 25

FOURTH EDITION

VITAMINS TO
ZONE REFINING

CONTENT INDEX (vol 25)

(with hyperlinks)

Edited by <http://www.watcherworld.narod.ru/>

VITAMINS	2
- SURVEY	3
- ASCORBIC ACID	10
- BIOTIN	27
- FOLIC ACID	36
- NIACIN, NICOTINAMIDE, AND NICOTINIC ACID	46
- PANTOTHENIC ACID	56
- PYRIDOXINE (B6)	66
- RIBOFLAVIN (B2)	74
- THIAMINE (B1)	84
- VITAMIN A	95
- VITAMIN B12	107
- VITAMIN D	124
- VITAMIN E	144
- VITAMIN K	151
WASTE TREATMENT, HAZARDOUS WASTE	159
WASTES, INDUSTRIAL	176
WATER	197
- SOURCES AND QUALITY ISSUES	198
- PROPERTIES	207
- POLLUTION	220
- ANALYSIS	230
- SUPPLY AND DESALINATION	235
- INDUSTRIAL WATER TREATMENT	258
- MUNICIPAL WATER TREATMENT	275
- SEWAGE	282
- REUSE	291
- TREATMENT OF SWIMMING POOLS, SPAS, AND HOT TUBS	295
- WATERPROOFING AND WATER/OIL REPELLENCY	306
WAXES	314
WOOD	320
WOOL	338
XANTHATES	359
X-RAY TECHNOLOGY	371
YEASTS	385
ZINC AND ZINC ALLOYS	396
ZINC COMPOUNDS	419
ZIRCONIUM AND ZIRCONIUM COMPOUNDS	426
ZONE REFINING	446

VISCOSE.

See FIBERS, REGENERATED CELLULOSES.

VISCOSITY.

See RHEOLOGICAL MEASUREMENTS.

VISCOSITY BREAKING, VISBREAKING.

See PETROLEUM, REFINERY PROCESSES, SURVEY

VITAMINS

Survey,
Ascorbic acid,
Biotin,
Folic acid,
Niacin, nicotinamide, and nicotinic acid,
Pantothenic acid,
Pyridoxine,
Riboflavin,
Thiamine,
Vitamin A,
Vitamin B₁₂,
Vitamin D,
Vitamin E,
Vitamin K,

SURVEY

Vitamins are specific organic compounds that are essential for normal metabolism. Many participate as cofactors or coenzymes in mammalian biochemical reactions. The common thread for the diverse chemical structures of the vitamins is that they are micronutrients. Micronutrients are compounds that are required in only small amounts and are not synthesized by humans, either at all or, at least, in sufficient quantity for metabolic needs. Vitamins are obtained from the diet or as synthetic preparations used in food fortification or supplements.

The term vitamin is derived from *vitamine*, a word created in 1911 (1) when it was believed that the antiberiberi factor (thiamine) was an amine essential to the maintenance of life (*vita*). Later discoveries showed that not all vitamins are amines. However, a form of the name has remained.

The history of the vitamins can be divided into five major, overlapping periods (2).

Prehistory to ca 1900

In this period, the empirical healing of certain diseases by foods was established. Examples (3) were the treatment of night blindness (vitamin A deficiency) with liver in many cultures over centuries, of beriberi (vitamin B₁ deficiency) by use of unpolished rice by the Japanese navy, of scurvy (vitamin C deficiency) by citrus fruits in the British navy or pine needle extracts by North American natives, and pellagra (niacin deficiency) by a dietary shift away from corn-based foods in many countries. Other, nondietary empirical treatments involved, eg, exposure of children in northern latitudes to sunlight to cure rickets (vitamin D deficiency) (4).

1880–1905

The ability to induce disease states in animals by manipulation of the diet was established in this period. The classical work by Eijkman (5), in which a beriberi-like condition was induced in chickens fed on polished rice, was significant. These findings led to the concept by Hopkins that small amounts of accessory growth factors are necessary for survival and growth.

1900–1972

During this 70 year period, all 13 of the substances now recognized as vitamins were discovered and isolated in pure form. Structure elucidation for each vitamin was completed, as was its total synthesis (Table 1).

Table 1. History of the Vitamins^a

Vitamin	Discovery	Isolation	Chemical structure	Synthesis
A	1909	1931	1931	1947
D	1918	1932	1936	1959
E	1922	1936	1938	1938
K	1929	1939	1939	1939
B ₁	1897	1926	1936	1936
B ₂	1920	1933	1935	1935
niacin	1936	1935	1937	1867
B ₆	1934	1938	1938	1939
B ₁₂	1926	1948	1956	1972
folic acid	1941	1941	1946	1946
pantothenic acid	1931	1938	1940	1940
biotin	1931	1935	1942	1943
C	1912	1928	1933	1933

^a Adapted from Ref. 4.

1933–Present

The first commercial synthesis of a vitamin occurred in 1933 when the Reichstein approach was employed to manufacture vitamin C (6). All 13 vitamins are available in commercial quantities, and their biological functions have largely been established (7). A list of Nobel prize winners associated with vitamin research is given in Table 2.

Table 2. Nobel Prizes for Vitamin Research^a

Year	Recipient	Field	Citation
1928	Adolf Windaus	chemistry	research into constitution of steroids and connection with vitamins
1929	Christiaan Eijkman	medicine, physiology	discovery of antineuritic vitamins
	Sir Frederick G. Hopkins	medicine, physiology	discovery of growth-stimulating vitamin
1934	George R. Minot, William P. Murphy, George H. Whipple	medicine, physiology	discoveries concerning liver therapy against anemias
1937	Sir Walter N. Haworth Paul Karrer	chemistry chemistry	research into constitution of carbohydrates and vitamin C research into constitution of carotenoids, flavins, and vitamins A and B ₂
	Albert Szent-Gyrgyi	medicine, physiology	discoveries in connection with biological combustion processes, with special reference to vitamin C and catalysis of fumaric acid

1938	Richard Kuhn	chemistry	work on carotenoids and vitamins
1943	Carl Peter Henrik Dam	medicine, physiology	discovery of vitamin K
	Edward A. Doisy	medicine, physiology	discovery of chemical nature of vitamin K
1953	Fritz A. Lipmann	medicine, physiology	discovery of coenzyme A and its importance for intermediary metabolism
1964	Konrad E. Bloch, Feodor Lynen	medicine, physiology	discoveries concerning mechanism and regulation of cholesterol and fatty acid metabolism
	Dorothy C. Hodgkin	chemistry	structural determination of vitamin B ₁₂
1967	Ragnar A. Granit	medicine, physiology	research which illuminated electrical properties of vision by studying wavelength discrimination by eye
	Halden K. Hartine	medicine, physiology	research on mechanisms of sight
	George Wald	medicine, physiology	research on chemical processes that allow pigments in eye retina to convert light into vision

^a From Ref. 2.

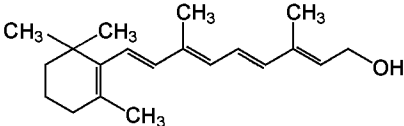
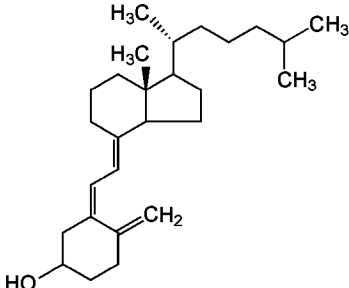
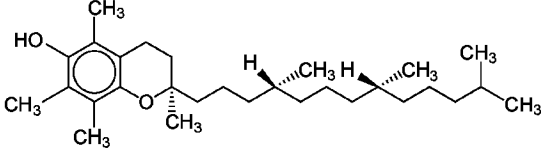
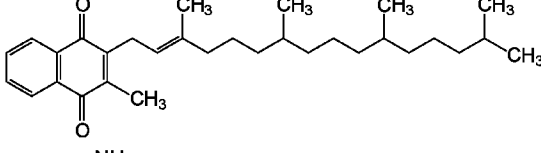
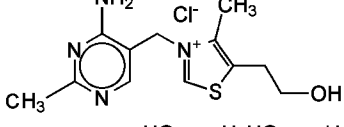
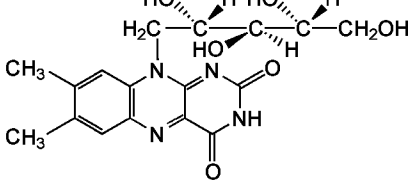
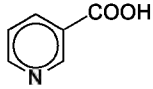
1955–Present

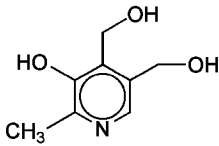
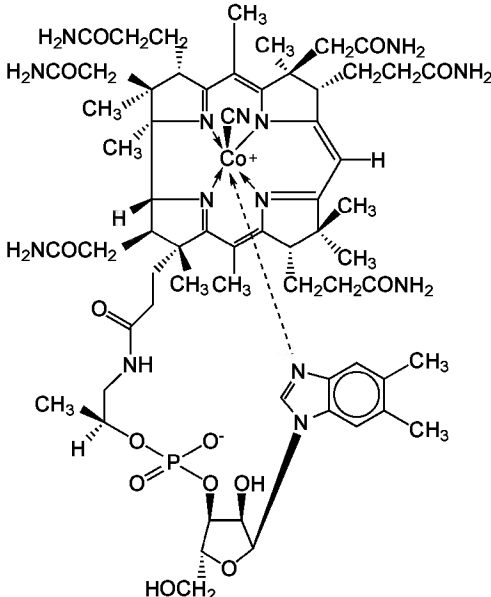
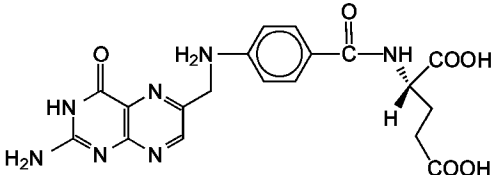
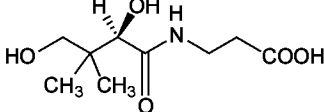
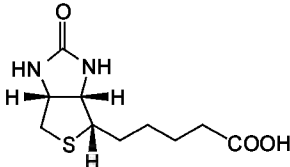
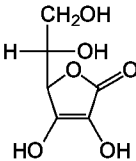
The possibility that vitamins might have physiological functions beyond the prevention of deficiency diseases was first recognized in 1955 with the finding (8) that niacin can affect serum cholesterol levels in humans. An explosion of research (9–11) in the intervening years has been aimed at establishing optimal vitamin levels and anticipating the health consequences.

Structure

Common names, chemical structure, and synonyms for the vitamins are given in Table 3. The names given to the vitamins and/or their letter designations do not follow a logical pattern and are of historical significance only. Despite this, the nomenclature is in common use.

Table 3. Vitamin Names and Structures

Vitamin	Structure	CAS Registry Number	Common name and synonyms
A		[68-26-8]	retinol, axerophthol
D		[520-91-2]	vitamin D ₃ , cholecalciferol, antirachitic vitamin
E		[59-02-9]	α-tocopherol, (RRR)-α-tocopherol, antisterility vitamin
K		[84-80-0]	vitamin K ₁ , phylloquinone, coagulation vitamin, antihemorrhagic vitamin, phytonadione
B ₁		[67-03-8]	thiamine, aneurin
B ₂		[83-88-5]	riboflavin, vitamin G, lactoflavin, hepatoflavin, ovoidflavin, verdoflavin
niacin		[59-67-6]	vitamin B ₃ , nicotinic acid

B ₆		[58-56-0]	pyridoxine, pyridoxol, antiacrodynia factor
B ₁₂		[68-19-9]	cobalamin, cyano-cobalamin
folic acid		[59-30-3]	folacin, pteroylglutamic acid, antianemia factor, fermentation <i>L. casei</i> factor
pantothenic acid		[79-83-4]	chick antidermatitis factor
biotin		[58-85-5]	<i>d</i> -biotin, vitamin H, anti-egg-white-injury factor, bios II, coenzyme R
C		[520-91-2]	ascorbic acid, antiscorbutic vitamin

Vitamins are classified by their solubility characteristics into fat-soluble and water-soluble groups. The fat-soluble vitamins A, E, D₃ and K result from the isoprenoid biosynthetic pathway. Vitamin A is derived by enzymic cleavage of the symmetrical C₄₀ beta-carotene, also known as pro-vitamin A. Vitamins E and K result from condensations of phytyldiphosphate (C₂₀) with aromatic components derived from shikimic acid. Vitamin D results from photochemical ring opening of 7-dehydrocholesterol, itself derived from squalene (C₃₀).

The water-soluble vitamins, B₁, B₂, B₆, B₁₂, niacin, folic acid, pantothenic acid, biotin, and C, have a much less common biosynthetic heritage. Furthermore, multiple biosynthetic paths exist for several of the vitamins. Thus vitamins B₁, B₂ and folic acid are derived from purines, with components coming from carbohydrates. Portions of vitamins B₁ and folic acid also originate in amino acids. One biosynthesis of niacin and that of vitamin B₆ are derived from glycerin. AcetylCoA serves as common precursor, in extremely complex processes, to biotin (via pimelic acid) and vitamin B₁₂ (through 5-aminolevulinic acid). Pantothenic acid is constructed from glyoxylic acid and beta-alanine, itself derived from several pathways. Vitamin C, a carbohydrate, is biosynthesized from glucose or galactose (4,12).

Vitamers

Although the 13 substances described in Table 3 are commonly recognized as the vitamins, significant complexities exist because of the following factors. (1) A number of similar chemical structures can exhibit, at least qualitatively, the biological activity of the vitamin. (2) The biological activity can vary by organism type, eg, in microorganisms vs humans. (3) The physiologically active form of the vitamin may be other than that indicated; for example, several water-soluble vitamins are active in co-enzyme form. (4) A metabolite may be the active form of the vitamin, as in the case of 1- α -25-dihydroxy-vitamin D₃. (5) Biochemical interconversion of several forms of the vitamins can occur. The best studied example is pyridoxine, pyridoxal, pyridoxamine, and their 5'-phosphates. (6) Synthetic commercial forms, often manufactured and sold because of their increased stability, are convertible in the target organism to the vitamin. Esters of α -tocopherol and retinol are the most prominent examples. These compounds are known as vitamers. Table 4 presents a partial list of these vitamers. Specifically segmented are the major commercial forms of the vitamin.

Table 4. Vitamers

Vitamin	Vitamers	Major commercial forms
A	retinal (vitamin A aldehyde), retinoic acid (vitamin A	vitamin A acetate, vitamin A propionate, vitamin A palmitate, vitamin A

D	acid), retinol cis-isomers, β -carotene, other carotenoids ergocalciferol (vitamin D ₂), 1-	acid, β -carotene, β -apo-8'-carotenal cholecalciferol (vitamin D ₃), ergocalciferol (vitamin D ₂)
E	α -25-dihydroxycholecalciferol (calcitriol) (RRR)- β -tocopherol, (RRR)- γ -tocopherol, (RRR)- δ -tocopherol, 2R- α -tocotri-enol, 2R- β -tocotrienol, 2R- γ -tocotrienol, 2R- δ -tocotrienol, all- <i>rac</i> - α -tocopherol	(RRR)- α -tocopherol, (RRR)- α -tocopheryl acetate, all- <i>rac</i> - α -tocopherol, all- <i>rac</i> - α -tocopheryl acetate, all- <i>rac</i> - α -tocopheryl succinate
K	vitamin K ₂ (menaquinone), vitamin K ₃ (menadione), 2-(R, <i>S</i>)-vitamin K ₁ (trans)	2-(R, <i>S</i>)-vitamin K ₁ (cis, trans-mixture), menadione sodium bisulfite, menadione dimethylpyrimidinol bisulfite
B ₁	thiamine diphosphate (cocarbox-ylase), thiamine ortho-phosphate	thiamine chloride, thiamine mononitrate, thiamine diphosphate, thiamine disulfide
B ₂	riboflavin mononucleotide (FMN), riboflavin dinucleo-tide (FAD)	riboflavin, riboflavin sodium phosphate
niacin	niacinamide (nicotinamide), nicotinamide adenine dinucleo-tide (NAD, coenzyme I), nicotinamide adenine dinucleotide phosphate (NADP, coenzyme II)	niacin, niacinamide
B ₆	pyridoxine-5'-phosphate, pyridoxal, pyridoxal-5'-phosphate (codecarboxylase), pyridoxamine, pyridoxamine-5'-phosphate	pyridoxine hydrochloride, pyridoxal-5'-phosphate
B ₁₂	methylcobalamin, adenosyl-cobalamin (coenzyme B ₁₂), hydroxocobalamin, aqua-cobalamin chloride	cyanocobalamin, hydroxocobalamin
folic acid	tetrahydrofolic acid, <i>N</i> -formyl-tetrahydrofolic acid, pterotic acid, pteroyltriglutamic acid, pteroylhexaglutamic acid	folic acid
pantothenic acid	coenzyme A, pantethin	calcium pantothenate, <i>d</i> -panthenol, <i>d,l</i> -panthenol
biotin	desthiobiotin	<i>d</i> -biotin
C	dehydroascorbic acid	ascorbic acid, ascorbyl palmitate, sodium ascorbate, calcium ascorbate, ascorbyl-2-polyphosphate

Properties

Because of their diverse structures, there are few common threads to vitamin chemical properties aside from their fat or water solubility. Of general concern in all applications, however, is vitamin stability. Table 5 provides generic information regarding stability under several conditions. Levels of stability vary greatly and are impacted by acid or base strength, light intensity, etc.

Table 5. Vitamin Stability^a

Vitamin	Oxygen	Light	100°C ^b	Acids ^b	Bases ^b
A	U	U	S	U	S
D	U	U	U	U	U
E	U	U	S	S	S
K	S	U	S	S	U
B ₁	U	U	U	S	U
B ₂	S	U	U	S	U
niacin	S	S	S	S	S
B ₆	S	U	U	S	S
B ₁₂	U	U	U	S	S
folic acid	U	U	U	U	S
pantothenic acid	S	S	U	U	U
biotin	S	S	U	S	S
C	U	U	U	S	U

^a S = stable; U = unstable.

^b In absence of oxygen.

The relative instability of many of the vitamins has led to the creation of derivatives and stabilized product forms capable of withstanding the rigors of food, animal feed, and pharmaceutical process conditions (see FEEDS AND FEED ADDITIVES; PHARMACEUTICALS). Depending on the vitamin and application, stabilized forms may involve spray drying in a suitable matrix (eg, gelatin), encapsulation (beadletting), or coating with fats or waxes (qv).

Analysis

Analytical methods for the vitamins involve physical and chemical, chromatographic, microbiological, and biological methods. The method of choice for a particular vitamin depends on the substance being measured, as well as its concentration. For example, detection of biotin, folic acid, and vitamin B₁₂ is accurately achieved by high pressure liquid chromatography (hplc) or gas-liquid chromatography (glc). However, chromatographic methods often are not sufficiently sensitive for the low levels at which these vitamins usually are found. Microbiological methods, though slower, more expensive, and usually less accurate, are then employed for these low level assays. Biological assays, although extremely significant in the history of the vitamins, are rarely used in the 1990s.

An excellent overview of vitamin analytics is available (13). Specifications for the vitamins are available for food use (14) and pharmaceutical use (15).

Physiological Function

Physiological functions as well as clinical symptoms that occur in humans deficient in specific vitamins are given in Table 6. It is becoming more authenticated that vitamins have additional potential health benefits when administered, via the diet or by supplementation, at levels above those required for obviating deficiency. Although for most vitamins the optimal levels are not yet established, some of the potential health benefits to be derived from vitamins are indicated (16). In one case, the level of scientific proof is such that the U.S. FDA has allowed "a health claim that women who are capable of

becoming pregnant and who consume adequate amounts of folate daily during their childbearing years may reduce their risk of having a pregnancy affected by spina bifida or other neural tube defects" (17).

Table 6. Vitamin Functions, Clinical Deficiency Symptoms, and Potential Health Benefits

Vitamin	Function	Clinical deficiency symptoms	Potential health benefits
A	needed for normal vision, reproduction, and maintenance of healthy skin, mucous membranes, bones, red blood cells, cell differentiation, and im-mune function	night blindness, xerophthalmia, loss of appetite, increased susceptibility to infections, skin disorders, poor growth, defective reproduction	regression of precancerous lesions, reduces measles-associated morbidity in children
D	enhances calcium absorption, regu-lates calcium and phosphorus metabolism, promotes minerali-zation of bones, plays role in cell differentiation	insufficient bone mineralization, skeletal deformities (rickets and osteomalacia)	protection against certain types of cancer, reduces risk of osteoporosis, psoriasis (topical)
E	lipid-soluble antioxidant, prevents lipid oxidation of membranes, needed for healthy blood cells and tissues, blocks nitrosamine formation, protects PUFAs from autoxidation, important for normal immune function	neuromuscular disorders, red blood cell rupture (both uncommon)	reduces risk of chronic disease (cardio-vascular, precancerous lesions, cancer), immunoenhancement, protec-tion from exercise-induced muscle injury, improves metabolic control, re-duces risk of complications in diabetes
K	essential for normal blood clotting, needed for formation and mainte-nance of healthy bones, essential for formation of osteocalcin (bone protein)	reduced blood clotting, increased risk for brittle bones	reduces risk of osteoporosis
B ₁	helps convert carbohydrates from food into energy, required by muscle, nervous system, and brain	mental confusion, loss of appetite, muscular weakness and paralysis, heart failure (beriberi)	important role in energy metabolism
B ₂	involved in carbohydrate, protein and fat metabolism	oral lesions, cracks in corners of mouth, dermatitis	important role in energy metabolism
niacin	involved in carbohydrate, protein, and fat metabolism; at pharma-cologic levels used to lower blood cholesterol	mouth soreness, diarrhea, abdominal pain, skin and nervous system disorders (pellagra)	reduces risk of atherosclerosis
B ₆	active in protein metabolism, required for bone health, involved in homocysteine metabolism	dermatitis, anemia, nervous system disorders	reduces risk of cardiovascular disease (also B ₁₂ and folic acid), immunoenhancement, reduces risk of osteoporosis, protection against carpal tunnel syndrome
folic acid	involved in amino acid interconversions, needed for red blood cell formation and DNA and RNA synthesis, protects against neural tube birth defects, involved in homocysteine metabolism	anemia, nervous system disorders, gastrointestinal lesions	reduces risk of cardiovascular disease (also B ₆ and B ₁₂), certain cancers, and precancerous lesions; birth defect prevention
pantothenic acid	active in carbohydrate, protein, and fat metabolism	abnormal sensations in legs and feet (rare)	
biotin	involved in protein, fat, and carbo-hydrate metabolism	dermatitis, hair loss	helps develop healthy nails
C	helps form and maintain collagen; important for healthy cartilage, bones, teeth, and gums; needed for wound healing; enhances nonheme iron absorption; water-soluble antioxidant protects cells from free-radical damage; blocks nitrosamine formation; increases body's resistance to infection	swollen or bleeding gums, joint pain, slow wound healing, lowered resistance to infection, scurvy	protection against chronic diseases (cancer, cardiovascular, eye disease), reduces symptoms of common cold, periodontal disease protection, improves fertility in men, reduces risk of osteoporosis

The Recommended Dietary Allowances (RDA) of the National Academy of Sciences are intended as daily intake guidelines for preventing deficiencies and maintaining reasonable reserves among most healthy people. These recommendations, revised every 5 to 10 years, are made by a committee of the U.S. Food and Nutrition Board and are based on scientific information that is available at the time of the review. This includes studies on subjects on deficient diets, nutrient balance studies, measures of tissue saturation, normal intakes of the nutrient, and extrapolation from animal data. The most recent RDAs are given in Table 7 (18). The 1989 RDA has a special requirement of 100 mg of vitamin C for cigarette smokers. Otherwise, there is no recognition of special needs for specific groups such as those using certain drugs chronically, people with disease, or the elderly (eg, over 65 years). Neither does it address the needs for the prevention of any of the chronic diseases.

Table 7. Recommended Dietary Allowances^a

Category	Age	Vitamin A, µg RE ^b	Vitami n D, µg ^c	Vitamin E, mg α-TE ^d	Vitami n K, µg	Thiami n, mg	Riboflavi n, mg	Niaci n, mg	Vitami n B ₆ , mg	Vitami n B ₁₂ , µg	Folat e, µg	Vitami n C, mg
infants	0,0–0.5	375	7.5	3	5	0.3	0.4	5	0.3	0.3	25	30
	0.5–1.0	375	10	4	10	0.4	0.5	6	0.6	0.5	35	35
children	1–3	400	10	6	15	0.7	0.8	9	1.0	0.7	50	40
	4–6	500	10	7	20	0.9	1.1	12	1.1	1.0	75	45
	7–10	700	10	7	30	1.0	1.2	13	1.4	1.4	100	45
males	11–14	1000	10	10	45	1.3	1.5	17	1.7	2.0	150	50

females	15–18	1000	10	10	65	1.5	1.8	20	2.0	2.0	200	60
	19–24	1000	10	10	70	1.5	1.7	19	2.0	2.0	200	60
	25–50	1000	5	10	80	1.5	1.7	19	2.0	2.0	200	60
	51 +	1000	5	10	80	1.2	1.4	15	2.0	2.0	200	60
	11–14	800	10	8	45	1.1	1.3	15	1.4	2.0	150	50
	15–18	800	10	8	55	1.1	1.3	15	1.5	2.0	180	60
	19–24	800	10	8	60	1.1	1.3	15	1.6	2.0	180	60
	25–50	800	5	8	65	1.1	1.3	15	1.6	2.0	180	60
pregnant lactating	51 +	800	5	8	65	1.0	1.2	13	1.6	2.0	180	60
		800	10	10	65	1.5	1.6	17	2.2	2.2	400	70
	1st six months		1000	10	12	65	1.6	1.8	20	2.1	2.6	280
2nd six months		1200	10	11	65	1.6	1.7	20	2.1	2.6	260	90

^a Ref. 18.

^b RE = retinol equivalents; 1 RE = 1 µg retinol or 6 µg β-carotene.

^c As cholecalciferol; 10 µg cholecalciferol = 400 IU vitamin D.

^d TE = α-Tocopherol equivalents; 1 α-TE = 1 mg (RRR)-α-tocopherol.

Because there is less information upon which to base dietary allowances for biotin and pantothenic acid, ranges of intake are provided, as in Table 8.

Table 8. Estimated Safe and Adequate Daily Dietary Intake ^a

Category	Age	Biotin, µg	Pantothenic acid, mg
infants	0–0.5	10	2
	0.5–1	15	3
	1–3	20	3
children and adolescents	4–6	25	3–4
	7–10	30	4–5
	11 +	30–100	4–7
adults		30–100	4–7

^a Ref. 18.

Along with increasing evidence of health benefits from consumption of vitamins at levels much higher than RDA recommendations comes concern over potential toxicity. This topic has been reviewed (19). Like all chemical substances, a toxic level does exist for each vitamin. Traditionally it has been assumed that all water-soluble vitamins are safe at any level of intake and all fat-soluble vitamins are toxic, especially at intakes more than 10 times the recommended allowances. These assumptions are now known to be incorrect. Very high doses of some water-soluble vitamins, especially niacin and vitamin B₃, are associated with adverse effects. In contrast, evidence indicates that some fat-soluble micronutrients, especially vitamin E, are safe at doses many times higher than recommended levels of intake. Chronic intakes above the RDA for vitamins A and D especially are to be avoided, however.

Production

Preparation of the vitamins in commercial quantities can involve isolation, chemical synthesis, fermentation, and mixed processes, including chemical and fermentation steps. The choice of process is economic, dictated by the need to obtain materials meeting specifications at the lowest cost. Current process technologies (ca 1997) employed for each vitamin are indicated in Table 9.

Table 9. Vitamin Production Processes

Isolation ^a	Chemical synthesis	Fermentation	Synthesis fermentation ⁿ
(vitamin A)	vitamin A	vitamin B ₂	vitamin B ₂
vitamin E ^b	vitamin D	vitamin B ₁₂	vitamin C
(vitamin D)	vitamin E ^c		
(vitamin K)	vitamin K		
	vitamin B ₁		
	niacin		
	vitamin B		
	folic acid		
	pantothenic acid		
	biotin		

^a Parentheses indicate minor processes primarily of historical significance.

^b (RRR)- α-Tocopherol.

^c All-*rac*-α-tocopherol.

More than one process is available for some of the vitamins. Further, manufacturers have developed variants of the classical syntheses during optimization. Whereas some of this information is available, as described in the individual sections on vitamins, much is closely held as trade secrets. Judging from the more recent patent literature, the assessment can be made that vitamin production technologies are in general mature. However, the economic value of these products drives continuing research aimed at breakthrough processes. Annual production of vitamins varies greatly, from ca 10 metric tons of vitamin B₁₂ to ca 50,000 metric tons of vitamin C.

Applications

It is generally assumed that adequate vitamin levels in humans can be obtained through a balanced diet. However, ongoing studies continue to indicate that the majority of the U.S. population is not receiving even the RDA through diet. Supplementary vitamins are thus provided for fortification of foods (20) and as oral or parenteral dosage forms.

The use of vitamins in humans consumes ca 40⁰ of vitamins made worldwide. The majority of the vitamins, particularly in countries outside the United States, are used in animal husbandry. It is well established (21) that vitamins are critical to animal productivity, especially under confined, rapid growth conditions. Newer information (22) has shown that vitamin E added to cattle feed has the additional effect of significantly prolonging beef shelf life in stores. Additional applications of vitamins exist. A small but growing market segment involves cosmetics (qv) (23). The use of the chemical properties of the vitamins, particularly as antioxidants (qv) in foods and, more recently, in plastics (vitamin E (24)), has emerged as a growing trend.

Economic Aspects

Vitamins are sold for direct application to foods and animal feeds. In addition, they are further processed into nutritional supplements. This last market is particularly significant in the United States. In many other countries, vitamins are regulated as drugs, leading to a much lower supplement usage.

It can be estimated that approximately \$3,000,000,000 of vitamins were sold in 1996. Market growth is slightly higher than population growth, but varies widely by individual vitamin, geographical area, and or application. The largest vitamin manufacturer is Hoffmann-La Roche. Other significant producers include BASF, Takeda, Eisai, and Rhøfne-Poulenc. Additional vitamins are produced in China, Russia and India.

Vitamin-Like Compounds

Although it is a general consensus that these 13 substances represent the known vitamins, other compounds upon which there is less agreement often enter the vitamin discussion (25). Included in this group are choline, myo-inositol (Bios I), vitamin F (essential fatty acids including linoleic acid, γ-linolenic acid, and arachidonic acid), α-lipoic acid, ubiquinones, plastoquinone, vitamin P (bioflavonoids), vitamin L (o-aminobenzoic acid), vitamin T (mixture including folic acid and vitamin B₁₂), vitamin U (L-methioninylmethylsulfonium chloride), orotic acid (vitamin B₁₃), and pyrroloquinoline quinone.

BIBLIOGRAPHY

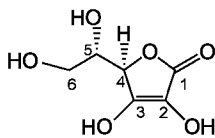
"Vitamins, Survey" in *ECT* 1st ed., Vol. 14, pp. 777–791, by H. R. Rosenberg and M. G. Weller, E. I. du Pont de Nemours & Co., Inc.; in *ECT* 2nd ed., Vol. 21, pp. 484–490, P. W. Fryth, Hoffmann-La Roche Inc.; in *ECT* 3rd ed., Vol. 24, pp. 1–7, P. W. Fryth, Hoffmann-La Roche Inc.

1. C. Funk, *J. Physiol. (London)* **43**, 395 (1911).
2. *Vitamins Basics*, F. Hoffmann-La Roche, Ltd., Basel, Switzerland, 1994, p. i.
3. K. Reese, *Today's Chem. Work*, **59** (Oct. 1995).
4. O. Isler and G. Brubacher, *Vitamine I: Fettlysliche Vitamine*, Georg Thieme Verlag, Stuttgart, Germany, 1982.
5. C. Eijkman, *Geweskund Tijdsche Ned. Indie* **30**, 295 (1896).
6. T. Reichstein and A. Grøssner, *Helv. Chim. Acta* **17**, 311 (1934).
7. L. J. Machlin, ed., *Handbook of Vitamins*, 2nd ed., Marcel Dekker, New York, 1991.
8. R. Altschul, A. Hoffer, and J. D. Stephen, *Arch. Biochem. Biophys.* **54**, 558 (1955).
9. H. E. Sauberlich and L. J. Machlin, eds., *Annal. NY Acad. Sci.* **669** (1992).
10. A. Bendich and C. E. Butterworth, Jr., *Micronutrients in Health and Disease Prevention*, Marcel Dekker, New York, 1991.
11. S. K. Gaby, A. Bendich, V. N. Singh, and L. J. Machlin, *Vitamin Intake and Health: A Scientific Review*, Marcel Dekker, New York, 1991.
12. O. Isler, G. Brubacher, S. Ghisla, and B. Krøutler, *Vitamine II: Wasserlysliche Vitamine*, Georg Thiem Verlag, Stuttgart, Germany, 1988.
13. J. Augustin, in the *Encyclopedia of Food Technology*, Vol. 4, John Wiley & Sons, Inc., New York, 1991, p. 2697.
14. Food and Nutrition Board, National Research Council, *Food Chemicals Codex*, 3rd ed., National Academy Press, Washington, D.C., 1981.
15. *The United States Pharmacopeia XXIII*(USP XXIII-NF XVIII), United States Pharmacopeial Convention, Inc., Rockville, Md., 1995.
16. V. Singh, personal communication, Hoffmann-La Roche Inc., Nutley, N.J., 1996.
17. *Fed. Reg.* **59**, 433 (Jan. 4, 1994).
18. Food and Nutrition Board, National Research Council, *Recommended Dietary Allowances*, 10th ed., National Academy Press, Washington, D.C., 1989.
19. *Vitamin Nutrition Information Service Backgrounduer*, Vol. 5 (No. 1), Hoffmann-La Roche Inc., Nutley, N.J., 1994.
20. J. Giese, *Food Technol.*, 110 (May 1995).
21. *Vitamin Compendium*, F. Hoffmann-La Roche & Co., Ltd., Basel, Switzerland, 1976.
22. J. A. Sherbeck, D. M. Wulf, J. B. Morgan, J. D. Tatum, G. C. Smith, and S. N. Williams, *J. Food Sci.* **60**, 250 (1995).
23. D. Djerassi, *The Role of Vitamins in Skin Care*, Hoffmann-La Roche Inc., Nutley, N.J., 1993.
24. S. F. Laermer, S. S. Young, and P. F. Zambetti, *Converting Mag.*, 80 (Feb. 1996).
25. M. M. Cody, in Ref. 7, p. 565.

John W. Scott
Hoffmann-La Roche Inc.

ASCORBIC ACID

Ascorbic Acid [50-81-1] (1) is the name recognized by the IUPAC-IUB Commission on Biochemical Nomenclature for Vitamin C (1). Other names are: L-ascorbic acid, L-xyloascorbic acid, and L-threo-hex-2-enoic acid γ -lactone. The name



(1) L-Ascorbic acid

implies the vitamin's antiscorbutic properties, eg, the prevention and treatment of scurvy. L-Ascorbic acid is widely distributed in plants and animals. The pure vitamin, C H₈O , mol wt 176.13, is a water-soluble, strongly reducing, optically active (chiral) white crystalline substance.

L-Ascorbic acid biosynthesis in plants and animals as well as the chemical synthesis starts from D-glucose. The vitamin and its main derivatives, sodium ascorbate, calcium ascorbate, and ascorbyl palmitate, are officially recognized by regulatory agencies and included in compendia such as the *United States Pharmacopeia/ National Formulary (USP/NF)* and the *Food Chemicals Codex (FCC)*.

The most significant chemical characteristic of L-ascorbic acid (1) is its oxidation to dehydro-L-ascorbic acid (L-threo-2,3-hexodiulosonic acid γ -lactone) (3) (Fig. 1). Vitamin C is a redox system containing at least three substances: L-ascorbic acid, monodehydro-L-ascorbic acid, and dehydro-L-ascorbic acid. Dehydro-L-ascorbic acid and the intermediate product of the oxidation, the monodehydro-L-ascorbic acid free radical (2), have antiscorbutic activity equal to L-ascorbic acid.

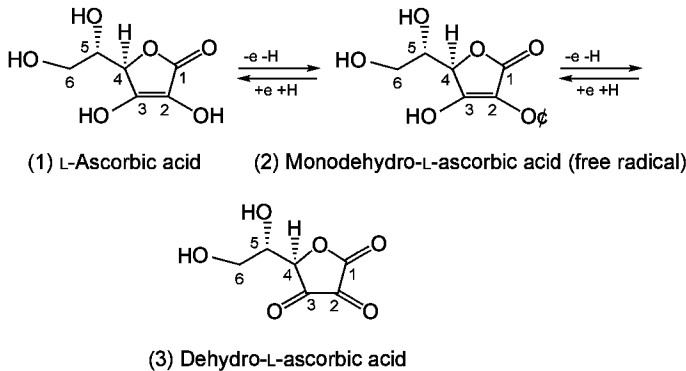


Fig. 1. Vitamin C redox system.

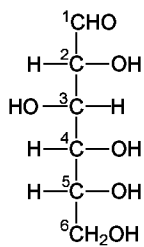
The reversible oxidation of L-ascorbic acid to dehydro-L-ascorbic acid is the basis for its known physiological activities, stabilities, and technical applications (2). The importance of vitamin C in nutrition and the maintenance of good health is well documented. Over 22,000 references relating only to L-ascorbic acid have appeared since 1966.

L-Ascorbic acid was isolated first by Albert Szent-Gyrgyi in 1930. However, its early history is associated with the etiology, treatment, and prevention of scurvy (3–5). Scurvy is one of the oldest diseases known to humankind. It affected many people in ancient times in Egypt, Greece, and Rome and influenced the course of history. Outbreaks of the disease resulting from inadequate amounts of vitamin C in rations spontaneously ended many military campaigns and long ocean voyages (6). During the Middle Ages, scurvy was endemic in northern Europe, occurring mostly in the winter season when fresh fruits and vegetables were unavailable. By the end of the seventeenth century, scurvy became a severe problem among sailors on long voyages. It was treated as a venereal disease, with disastrous results.

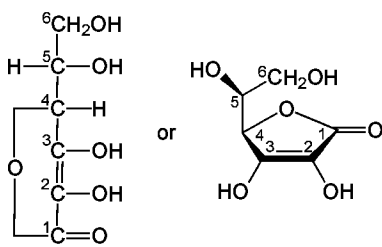
The first clues to the treatment of scurvy occurred in 1535–1536 when Jacques Cartier, on advice from Newfoundland Indians, fed his crew an extract from spruce tree needles to cure an epidemic. Various physicians were recommending the use of citrus fruits to cure scurvy in the mid-sixteenth century. Two hundred years later, in 1753, it was proved by Dr. James Lind, in his famous clinical experiment, that scurvy was associated with diet and caused by lack of fresh vegetables. He also demonstrated that oranges and lemons were the most effective cure against this disease. In 1753, in *A Treatise on the Scurvy*, Lind published his results and recommendations (7). Forty-two years later, in 1795, the British Navy included lemon juice in seamen's diets, resulting in the familiar nickname "limeys" for British seamen. Evidence has shown that even with undefined scorbutic symptoms, vitamin C levels can be low, and can cause marked diminution in resistance to infections and slow healing of wounds.

Research leading to the discovery of vitamin C began in 1907 when it was observed by Axel Holst and Theodor Fröhlich that guinea pigs were as susceptible to scurvy as humans and that the disease could be produced experimentally in these animals (8). These findings led to the development of an assay for the biological determination of antiscorbutic activity of food products (9).

Between 1910 and 1921, the vitamin was obtained in almost pure form from lemons and some of the physical and chemical properties were determined (10). It was discovered that vitamin C is easily destroyed by oxidation and best protected by reducing agents, and that 2,6-dichlorophenolindophenol is reduced by solutions of the vitamin. Subsequent studies showed that the dichlorophenolindophenol test measured only the vitamin and not dehydroascorbic acid, which also has antiscorbutic activity. The isolation of vitamin C was first accomplished by Szent-Gyrgyi in 1928 from cabbage, paprika, and the adrenal glands of animals (11). The relationship between vitamin C and the antiscorbutic factor was established in 1932 by Szent-Gyrgyi and, at the same time, by King and Waugh (12). Also in 1932, the chemical structure of ascorbic acid was determined independently by Haworth and Hirst (13). One year later, in 1933, Reichstein synthesized D-ascorbic acid and L-ascorbic acid (14). L-Ascorbic acid was synthesized from D-glucose (4) because the chiral centers of C-2 and C-3 were in the correct configuration to become C-4 and C-5, respectively, of L-ascorbic acid (15).



(4) D-Glucose

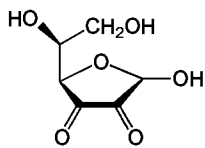


(1) L-Ascorbic acid

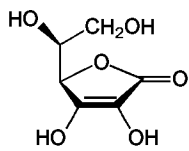
This synthesis was the first step toward industrial vitamin production, which began in 1936. The synthetic product was shown to have the same biological activity as the natural substance. It is reversibly oxidized in the body to dehydro-L-ascorbic acid (3) (*L-threo*-2,3-hexodiulosonic acid γ -lactone), a potent antiscorbutic agent with full vitamin activity. In 1937, Haworth and Szent-Gyurgyi received the Nobel Prize for their work on vitamin C.

Structure Determination

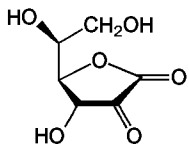
Albert Szent-Gyurgyi demonstrated that "hexuronic acid," C₆H₈O₇, which was first isolated from sources such as cabbage juice, was identical to vitamin C (*L*-ascorbic acid). Ultraviolet absorption studies carried out in the mid-1930s led to the proposal that *L*-ascorbic acid was 3-oxo-*L*-sorbose (5), which exists in a variety of tautomeric forms, including (6), (7), and *L*-ascorbic acid (1). The chemical structure of *L*-ascorbic acid was determined independently by Haworth and Hirst using x-ray crystallographic techniques (13). Soon afterwards, additional data supporting the structure 1 were published by other authors (16–18). Excellent review articles summarizing the early work on vitamin C have appeared (19,20). Years later, with improved x-ray techniques (21,22) and neutron-diffraction methods (23), the structure of *L*-ascorbic acid was refined (24,25). The five-membered ring containing the enediol group is almost planar. The conformation of the side chain in the crystal is as shown by structure 1. The chiral center at position C-4 has the *R*- or *D*-configuration, whereas C-5 has the *S*- or *L*-configuration.



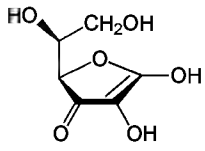
(5) 3-oxo-L-sorbose



(1)



(6)



(7)

As a result of having two chiral centers, four stereoisomers of ascorbic acid are possible (Table 1) (Fig. 2). Besides *L*-ascorbic acid (Activity = 1), only *D*-araboascorbic acid (erythorbic acid (9)) shows vitamin C activity (Activity = 0.025–0.05). The *L*-ascorbic acid structure (1) in solution and the solid state are almost identical. Ascorbic acid crystallizes in the space group P2₁ with four molecules in the unit cell. The crystal data are summarized in Table 2.

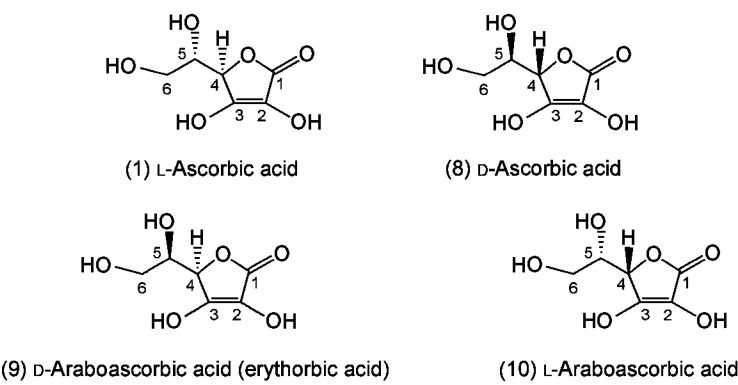


Fig. 2. Isomers of ascorbic acid.

Table 1. Isomeric Ascorbic Acids

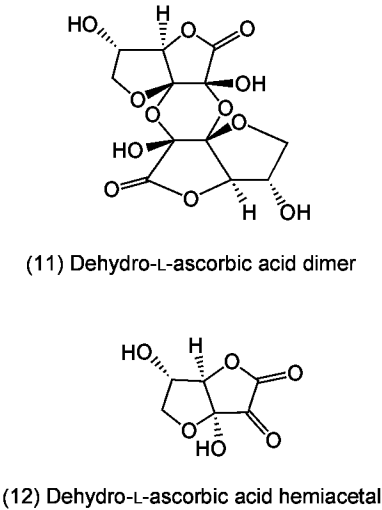
Substance	Structure	Mp, °C	[α] _D (H ₂ O), °	Activity
L-ascorbic acid	(1)	192	+24	1
D-xyloascorbic acid (D-ascorbic acid, D- <i>threo</i> -hex-2-enonic acid γ-lactone)	(8)	192	23	0
D-araboascorbic acid (erythorbic acid, D- <i>erythro</i> -hex-2-enonic acid γ-lactone)	(9)	174	18.5	0.025–0.05
L-araboascorbic acid (L- <i>erythro</i> -hex-2-enonic acid γ-lactone)	(10)	170	+17	0

Table 2. X-Ray Crystal Data of L-Ascorbic Acid^a

Space group	<i>P</i> 2 ₁
<i>a</i>	1.7299 nm
<i>b</i>	0.6353 nm
<i>c</i>	0.6411 nm
β	102° 11′
<i>V</i>	0.68859 nm ³
<i>Z</i>	4
<i>d</i> _{calcd}	1.699 g cm ^{−3}
μ(CuK _α)	13.9 cm ^{−1}

^a Ref. 24.

Crystalline dehydro-L-ascorbic acid (3) is reported to exist as the dimer (11) (26). In water, it is present as the hemiacetal monomer (12).



Properties

Physical Properties. Table 3 contains a summary of the physical properties of L-ascorbic acid. Properties relating to the structure of vitamin C have been reviewed and summarized (32). Stabilization of the molecule is a consequence of delocalization of the π -electrons over the conjugated enediol system. The highly acidic nature of the H-atom on C-3 has been confirmed by neutron diffraction studies (23).

Table 3. Physical Properties of L-Ascorbic Acid

Property	Characteristic(s)	Reference
appearance	white, odorless, crystalline solid with a sharp acidic taste	
formula; mol. wt.	C H ₈ O ; 176.13	

crystalline form	monoclinic; usually plates, sometimes needles	
mp °C	190–192 (dec)	27
density, g cm ³	1.65	27
optical rotation	[α] ²⁵ +20.5° to +21.5° (c = 1 in water) [α] ²³ +48 (c = 1 in methanol)	27
pH		27
5 mg mL	3	
50 mg mL	2	
p <i>K</i> ₁	4.17	27
p <i>K</i> ₂	11.57	27
redox potential	first stage: E = 0.166 V (pH 4)	27
solubility, g mL		27
water	0.33	
95 wt. % ethanol	0.033	
absolute ethanol	0.02	
glycerol USP	0.01	
propylene glycol	0.05	
ether	insoluble	
chloroform	insoluble	
benzene	insoluble	
petroleum ether	insoluble	
oils	insoluble	
fats	insoluble	
fat solvents	insoluble	
	Spectral Properties	
uv	pH 2: E _{max} (1° o, 1 cm) 695 at 245 nm (nondissociated form) pH 6.4: E _{max} (1° o, 1 cm) 940 at 265 nm (monodissociated form)	28
ir (KBr)	characteristic wavelengths, cm ⁻¹ 3455, 3405, 3155 v OH groups 2570 associated OH groups 1770, 1670 carbonyl lactone 1254 C–O–C lactone 1057 δ OH groups	29
nmr ^a	¹ H nmr (D ₂ O) δ 4.97 (d, 1 H, J _{4,5} = 2 Hz, H-4), 4.10 (ddd, 1 H, J _{5,6} = 5 and 7 Hz, J _{4,5} = 2 Hz, H-5), 3.78 (m, 2 H, C-6) ¹³ C nmr (D ₂ O) δ 174.0 (C-1), 118.8 (C-2), 156.3 (C-3) 77.1 (d, J _{C-4,H-4} = 158 Hz, C-4) 69.9 (d, J _{C-5,H-5} = 145.0 Hz, C-5), 63.2 (t, J _{C-6,H-6} = 145.0 Hz, C-6)	30 31

^a d = Doublet; ddd = doublet of doublet of doublet ; m = multiplet; and t = triplet.

Chemical Properties. The most significant chemical property of L-ascorbic acid is its reversible oxidation to dehydro-L-ascorbic acid. Dehydro-L-ascorbic acid has been prepared by uv irradiation and by oxidation with air and charcoal, halogens, ferric chloride, hydrogen peroxide, 2,6-dichlorophenolindophenol, neutral potassium permanganate, selenium oxide, and many other compounds. Dehydro-L-ascorbic acid has been reduced to L-ascorbic acid by hydrogen iodide, hydrogen sulfide, 1,4-dithiothreitol (1,4-dimercapto-2,3-butanediol), and the like (33).

The degradation of L-ascorbic acid in aqueous solution depends on several factors, eg, pH, temperature, presence of oxygen, or metals. Comprehensive reviews of degradation reactions and mechanisms have been published (34–37). In aqueous solution, L-ascorbic acid is more sensitive in the presence of oxygen to bases than to acids. The pH range with the highest stability is between 4 and 6. L-Ascorbic acid is sensitive to heat. In the presence of oxygen and heat, it is oxidized at a rate proportional to the temperature rise. On standing, dehydro-L-ascorbic acid (3), the initial oxidation product, undergoes irreversible hydrolysis to 2,3-dioxo-L-gulonic acid (*threo*-2,3-hexodiulosonic acid (13)), which is further oxidized to oxalic acid (14) and L-threonic acid ((*R*-(*R**,*S**)-2,3,4-trihydroxybutanoic acid) (15) (Fig. 3).

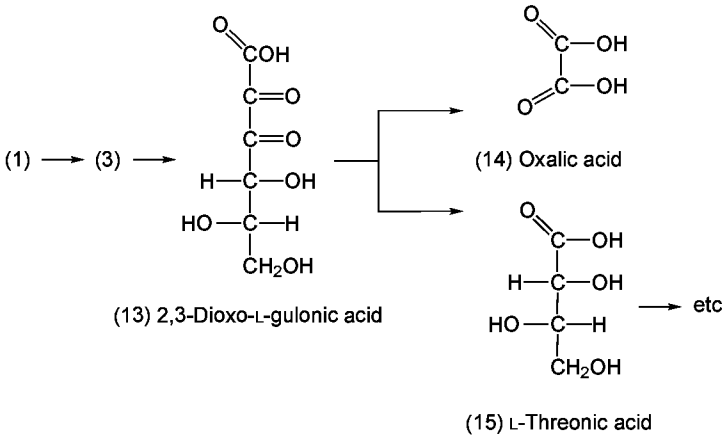


Fig. 3. Degradation of L-ascorbic acid.

In acidic solution, the degradation results in the formation of furfural, furfuryl alcohol, 2-furoic acid, 3-hydroxyfurfural, furoin, 2-methyl-3,8-dihydroxychroman, ethylglyoxal, and several condensation products (36). Many metals, especially copper, catalyze the oxidation of L-ascorbic acid. Oxalic acid and copper form a chelate complex which prevents the ascorbic acid-copper-complex formation and therefore oxalic acid inhibits effectively the oxidation of L-ascorbic acid. L-Ascorbic acid can also be stabilized with metaphosphoric acid, amino acids, 8-hydroxyquinoline, glycols, sugars, and trichloroacetic acid (38). Another catalytic reaction which accounts for loss of L-ascorbic acid occurs with enzymes, eg, L-ascorbic acid oxidase, a copper protein-containing enzyme.

Stability. Ascorbic acid, a white crystalline compound, is very soluble in water and has a sharp, acidic taste. In solution, the vitamin oxidizes on exposure to air, light, and elevated temperatures. Solutions of ascorbic acid turn yellowish, followed by development of a tan color. Ascorbic acid is stable to air when dry but gradually darkens on exposure to light.

Synthesis

Chemical Synthesis. The first synthesis of ascorbic acid was reported in 1933 by Reichstein and co-workers (14,39–42) (Fig. 4). Similar, independent reports published by Haworth and co-workers followed shortly after this work (13,43–45). L-Xylose (16) was converted by way of its osazone (17) into L-xylosone (18), which reacted with hydrogen cyanide forming L-xylonitrile (19). L-Xylonitrile cyclized under mild conditions to the cycloimine of L-ascorbic acid. Hydrolysis of the cycloimine yielded L-ascorbic acid. The yield for the conversion of L-xylosone to L-ascorbic acid was ca 40%.

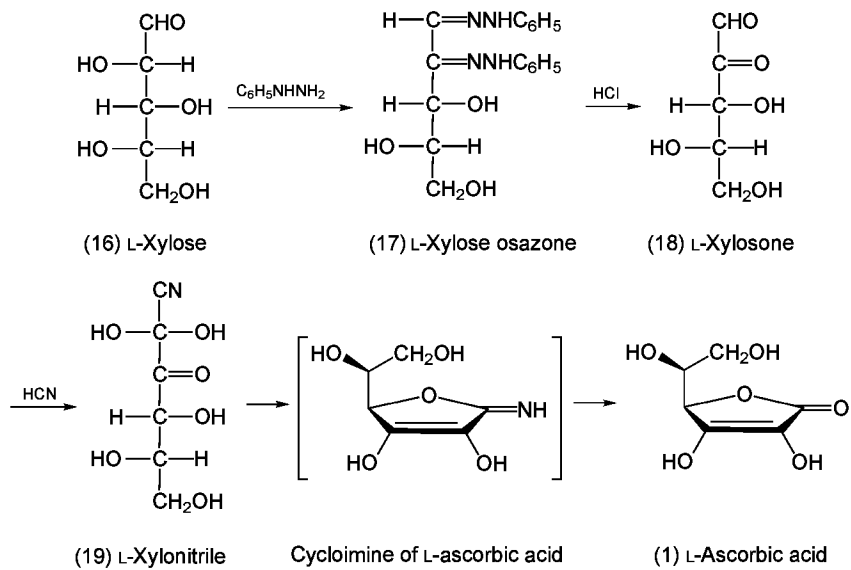


Fig. 4. First syntheses of L-ascorbic acid.

The L-xylosone pathway to L-ascorbic acid was never commercialized because L-xylosone was not readily available and was too expensive to prepare. The route, however, was valuable for L-ascorbic acid structure determination and for the preparation of derivatives.

Most current industrial vitamin C production is based on the efficient second synthesis developed by Reichstein and Grüssner in 1934 (15). Various attempts to develop a superior, more economical L-ascorbic acid process have been reported since 1934. These approaches, which have met with little success, are summarized in Crawford's comprehensive review (46). Currently, all chemical syntheses of vitamin C involve modifications of the Reichstein and Grüssner approach (Fig. 5). In the first step, D-glucose (4) is catalytically (Ni-catalyst) hydrogenated to D-sorbitol (20). Oxidation to L-sorbose (21) occurs microbiologically with *Acetobacter suboxydans*. The isolated L-sorbose is reacted with acetone and sulfuric acid to yield 2,3:4,6 diacetone-L-sorbose, (DAS) (2,3:4,6-bis-O-isopropylidene-α-L-sorbofuranose) (22). The remaining unprotected primary hydroxyl group of DAS is oxidized, either catalytically (O_2 , Pd), electrochemically, or by sodium hypochlorite with nickel chloride as catalyst to the corresponding carboxylic acid. The resulting 2,3:4,6-diacetone-2-keto-L-gulonic acid (DAG) (2,3:4,6-bis-O-isopropylidene-2-oxo-L-gulonic acid) (23) is treated with hydrogen chloride in an inert solvent system. Under these conditions, acetone removal occurs followed by consecutive lactonization and enolization to form L-ascorbic acid (1).

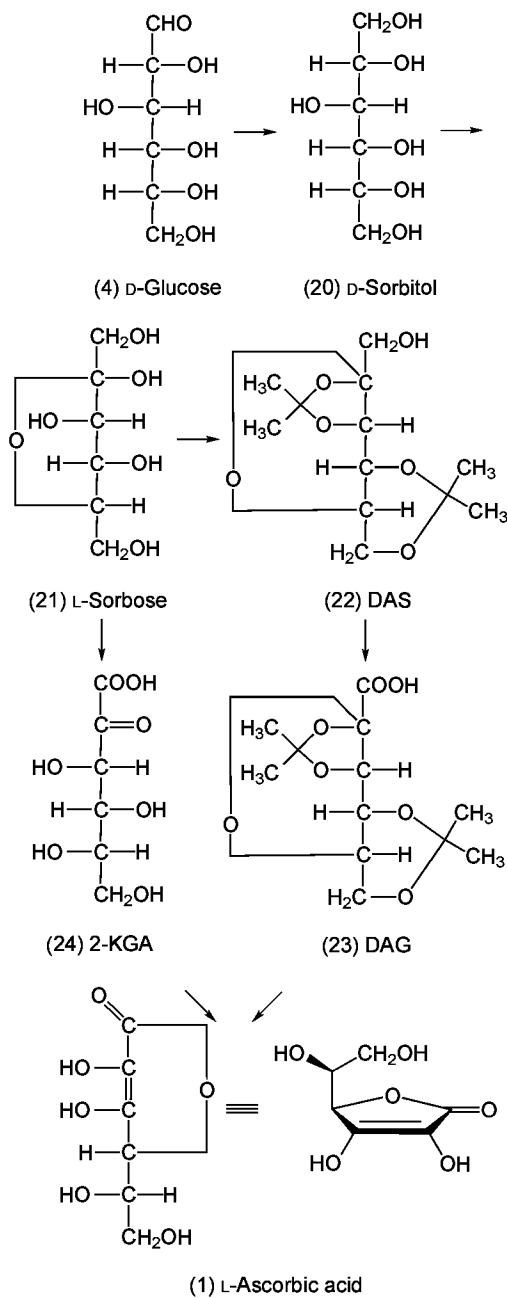


Fig. 5. Syntheses of L-ascorbic acid.

Fermentation. Much time and effort has been spent in undertaking to find fermentation processes for vitamin C (47). One such approach is now practiced on an industrial scale, primarily in China. It is not certain, however, whether these processes will ultimately supplant the optimized Reichstein synthesis. One important problem is the instability of ascorbic acid in water in the presence of oxygen; it is thus highly unlikely that direct fermentation to ascorbic acid will be economically viable. The successful approaches to date involve fermentative preparation of an intermediate, which is then converted chemically to ascorbic acid.

L-Sorbose to 2-KGA Fermentation. In China, a variant of the Reichstein-Grössner synthesis has been developed on an industrial scale (see Fig. 5). L-Sorbose is oxidized directly to 2-ketogulonic acid (2-KGA) (24) in a mixed culture fermentation step (48). Acid-catalyzed lactonization and enolization of 2-KGA produces L-ascorbic acid (1).

A Chinese publication (47) with 17 references reviews the use of genetically engineered microorganisms for the production of L-ascorbic acid and its precursor, 2-KGA (49). For example, a 2-keto-L-gulonic acid fermentation process from sorbose has been published with reported yields over 80% (50).

D-Glucose to 2-KGA Fermentation. A different fermentative route to 2-KGA proceeds via 2,5-diketo-L-gluconic acid (51,52). In a two-stage fermentation (Shionogi-Process), D-glucose (4) is oxidized to 2,5-diketo-D-gluconic acid (2,5-DKGA) (25) with *Erwinia* sp., followed by stereospecific reduction at C-5 by *Corynebacter* sp., forming 2-ketogulonic acid (24) (Fig. 6). The 2-KGA is rearranged upon treatment with acid to give ascorbic acid. A production of ascorbic acid by this route is reportedly being developed, ca 1997. An analogous process using a recombinant strain of *Erwinia citreus* was developed at Biogen (53).

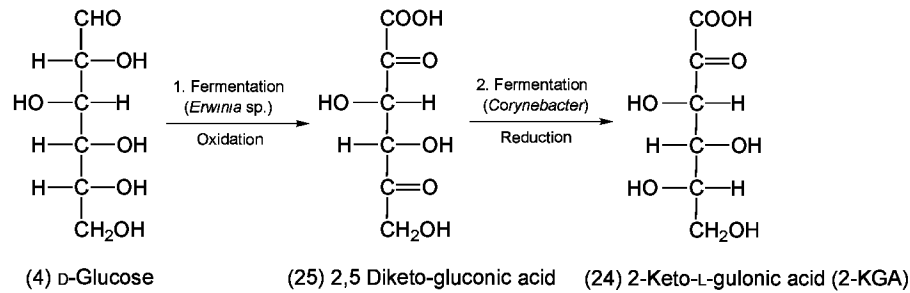


Fig. 6. D-Glucose to 2-KGA fermentation.

D-Glucose to L-Ascorbic Acid Fermentation. The direct heterotrophic fermentation of D-glucose to L-ascorbic acid with algae is

disclosed in a European Patent Application (54). The overall yield of L-ascorbic acid is 4° o.

Industrial Technology

Vitamin C was the first vitamin to be manufactured by chemical synthesis on an industrial scale. Major suppliers of vitamin C are Hoffmann-La Roche, BASF, Takeda, E. Merck, and various companies in China. Additional production occurs in Eastern Europe and India.

Reichstein and Grössner's second L-ascorbic acid synthesis became the basis for the industrial vitamin C production. Many chemical and technical modifications have improved the efficiency of each step, enabling this multistep synthesis to remain the principal, most economical process up to the present (ca 1997) (46). L-Ascorbic acid is produced in large, integrated, automated facilities, involving both continuous and batch operations. The process steps are outlined in Figure 7. Procedures require ca 1.7-kg L-sorbose kg of L-ascorbic acid with ca 66° o overall yield in 1977 (55). Since 1977, further continuous improvement of each vitamin C production step has taken place. Today's overall ascorbic acid yield from L-sorbose is ca 75° o. In the mid-1930s, the overall yield from L-sorbose was ca 30° o.

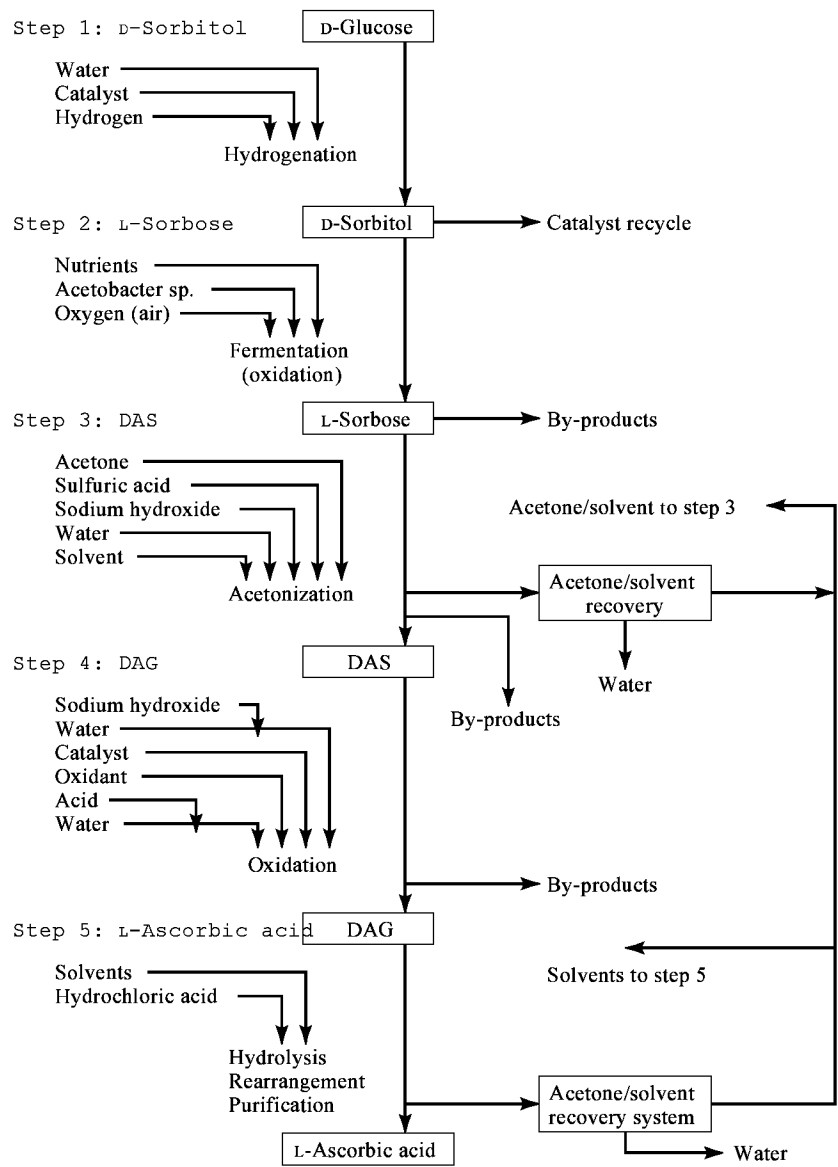


Fig. 7. L-Ascorbic acid manufactured by Reichstein and Grössner Synthesis.

The catalytic hydrogenation of D-glucose to D-sorbitol is carried out at elevated temperature and pressure with hydrogen in the presence of nickel catalysts, in both batch and continuous operations, with >97% yield (56,57). The cathodic reduction of D-glucose to L-sorbitol has been practiced (58). D-Mannitol is a by-product (59).

Sterile aqueous D-sorbitol solutions are fermented with *Acetobacter suboxydans* in the presence of large amounts of air to complete the microbiological oxidation. The L-sorbose is isolated by crystallization, filtration, and drying. Various methods for the fermentation of D-sorbitol have been reviewed (60). *Acetobacter suboxydans* is the organism of choice as it gives L-sorbose in >90% yield (61). Large-scale fermentations can be carried out in either batch or continuous modes. In either case, sterility is important to prevent contamination, with subsequent loss of product.

In the third step, L-sorbose is reacted with acetone and excess sulfuric acid at low temperatures. The sorbose dissolves on conversion into the 2,3-mono-O-isopropylidene-L-sorbose (2,3 monoacetone-L-sorbose) (MAS), and 2,3:4,6-bis-O-isopropylidene-α-L-sorbofuranose (2,3:4,6-diacetone-L-sorbose) (DAS). The equilibrium mixture consists of ca 65° o DAS and 35° o MAS. The sulfuric acid acts as a catalyst and a dehydrating agent. The reaction mixture is worked up by dilution, neutralization, and extraction to separate the DAS from the MAS. MAS is recycled and the mother liquors are distilled to recover acetone and solvents. The original Reichstein and Grössner process has been optimized over decades by other workers, giving ca 85° o yield (55). Other acidic catalysts, besides sulfuric acid, have been investigated, eg, zinc chloride–phosphoric acid, *p*-toluenesulfonic acid, copper(II) sulfate, and cationic exchange resins. Ferric chloride and perchloric acid are active catalysts that give >90% yields of DAS with azeotropic removal of the water (62). Other methods, including use of acetone dimethylacetal as a water removal reagent, have been reported (63). The use of other ketonic and aldehydic protective agents for L-sorbose, eg, cyclohexanone and its derivatives, formaldehyde, benzaldehyde, etc, has also been described (15,64).

2,3:4,6-Diacetone-L-sorbose (DAS) is oxidized at elevated temperatures in dilute sodium hydroxide in the presence of a catalyst (nickel chloride for bleach or palladium on carbon for air) or by electrolytic methods. After completion of the reaction, the mixture is worked up by acidification to 2,3:4,6-bis-O-isopropylidene-2-oxo-L-gulonic acid (2,3:4,6-diacetone-2-keto-L-gulonic acid) (DAG), which is isolated through filtration, washing, and drying. With sodium hypochlorite–nickel chloride, the reported DAG yields are >90% (65). The oxidation with air has been reported, and a practical process was developed with palladium–carbon or platinum–carbon as catalyst (66,67). The electrolytic oxidation with nickel salts as the catalyst has also

been published (68).

DAG is treated with ethanol and hydrochloric acid in the presence of inert solvent, eg, chlorinated solvents, hydrocarbons, ketones, etc. The L-ascorbic acid precipitates from the mixture as it forms, minimizing its decomposition (69). Crude L-ascorbic acid is isolated through filtration and purified by recrystallization from water. The pure L-ascorbic acid is isolated, washed with ethanol, and dried. The mother liquor from the recrystallization step is treated in the usual manner to recover the L-ascorbic acid and ethanol contained in it. The crude L-ascorbic acid mother liquor contains solvents and acetone liberated in the DAG hydrolysis. The solvents are recovered by fractional distillation and recycled. Many solvent systems have been reported for the acid-catalyzed conversion of DAG to L-ascorbic acid (46). Rearrangement solvent systems are used which contain only the necessary amount of water required to give >80% yields of high purity crude L-ascorbic acid (70).

The DAG conversion to L-ascorbic acid also can occur by a base-catalyzed mechanism. Methyl 2-oxo-L-gulonate (methyl DAG) is converted, on treatment with sodium methoxide, to sodium-L-ascorbate, which is then acidified to L-ascorbic acid. Various solvent systems have been evaluated and reported (46).

Packaging

L-Ascorbic acid is screened or pulverized into a variety of particle sizes. It is usually packaged in 25-kg and 50-kg quantities in standard, polyethylene-lined containers, eg, fiber drums, corrugated boxes, etc. The recommended storage conditions are low humidity and temperatures of <23°C.

Environmental Issues

The environmental concerns of an ascorbic acid manufacturing facility are those typical of a chemical processing plant. Its operating design must be patterned to conform to environmental protection regulations. Measures must be taken to contain solvents and to keep emissions within official guidelines. Special condensers, continuous instrumental monitoring, and emergency containment and cleanup systems are required. Wastewater-treatment facilities have to be provided to remove by-product organics and inorganics from effluent streams before disposal. The extent of these treatment facilities depends upon the location of the plant and the local tolerances. Usually, there are secondary treatment facilities for organic removal; at some plant sites, additional treatment may be required to remove inorganic salts.

Economic Aspects

Strong growth in demand during the period 1985–95 from existing and new applications led to firm pricing and expansion of world capacity. Total world demand in 1995 was estimated to be approximately 60,000 metric tons. Balance of capacity and demand at that time resulted in a selling price of \$16.00/kg. More recently (ca 1997), prices have dropped to less than \$10/kg.

Continuing demand from food, pharmaceutical, animal nutrition, and cosmetic applications will result in growth rates of 3–4% annually throughout the remainder of the 1990s.

Specifications

Specifications for ascorbic acid, sodium ascorbate, calcium ascorbate, and ascorbyl palmitate are found in the *United States Pharmacopeia/National Formulary* (71) and the *Food Chemicals Codex* (72). The official assay for all four compounds is the iodimetric titration with 0.1 N iodine solution and starch as the indicator.

Analytical Methods

Many different analytical methods have been developed to determine L-ascorbic acid in feed, biological, and pharmaceutical samples. An excellent review article describes the methodology for finding the proper L-ascorbic acid assay method (73). Comprehensive reviews of all analytical methods, including the extraction of ascorbic acid from foods and biological samples, have been published (74). Ascorbic acid has been determined by a variety of methods, including uv absorption; redox and derivatization reactions; electrochemical and enzymatic oxidation reactions; chromatographic, eg, hplc methods; and biological methods with animals. The guinea pig, one of the few animal species requiring ascorbic acid, is used in the bioassay for ascorbic acid activity. The various methods practiced have been described in detail (75).

Because of the time and expense involved, biological assays are used primarily for research purposes. The first chemical method for assaying L-ascorbic acid was the titration with 2,6-dichlorophenolindophenol solution (76). This method is not applicable in the presence of a variety of interfering substances, eg, reduced metal ions, sulfites, tannins, or colored dyes. This 2,6-dichlorophenolindophenol method and other chemical and physiochemical methods are based on the reducing character of L-ascorbic acid (77). Colorimetric reactions with metal ions as well as other redox systems, eg, potassium hexacyanoferrate(III), methylene blue, chloramine, etc, have been used for the assay, but they are unspecific because of interferences from a large number of reducing substances contained in foods and natural products (78). These methods have been used extensively in fish research (79). A specific photometric method for the assay of vitamin C in biological samples is based on the oxidation of ascorbic acid to dehydroascorbic acid with 2,4-dinitrophenylhydrazine (80). In the microfluorometric method, ascorbic acid is oxidized to dehydroascorbic acid in the presence of charcoal. The oxidized form is reacted with *o*-phenylenediamine to produce a fluorescent compound that is detected with an excitation maximum of ca 350 nm and an emission maximum of ca 430 nm (81).

Another method that determines both ascorbic acid and dehydroascorbic acid first reduced the dehydroascorbic acid to ascorbic acid and then retains the ascorbic acid on an anionic Sephadex column (82). The ascorbic acid is oxidized on the column to dehydroascorbic acid by *p*-benzoquinone, which simultaneously elutes the dehydroascorbic acid. The dehydroascorbic acid is reacted with 4-nitro-1,2-phenylenediamine and absorbance of the resulting yellow solution produced is measured at 375 nm.

Chromatographic methods, notably hplc, are available for the simultaneous determination of ascorbic acid as well as dehydroascorbic acid. Some of these methods result in the separation of ascorbic acid from its isomers, eg, erythorbic acid and oxidation products such as diketogulonic acid. Detection has been by fluorescence, uv absorption, or electrochemical methods (83–85). Polarographic methods have been used because of their accuracy and their ease of operation. Ion exclusion (86) and ion suppression (87) chromatography methods have recently been reported. Other methods for ascorbic acid determination include enzymatic, spectroscopic, paper, thin layer, and gas chromatographic methods. Excellent reviews of these methods have been published (73,88,89).

Uses

L-Ascorbic acid is used as a micronutrient additive in pharmaceutical, food, feed, and beverage products, as well as in cosmetic applications. The over-the-counter (OTC) vitamin market is strong, growing in demand, and vitamin C is available in the form of pills and tablets to supplement the daily diet to maintain peak physical performance.

Industrial uses of L-ascorbic acid relate to its antioxidant and reducing properties. It is used as an antioxidant in the commercial preparation of beer, fruit juices, cereals, and canned and frozen foods, etc.

A proposal was made to use L-ascorbic acid as an antioxidant replacement for sulfites in the food industries (90,91). Ascorbic acid's antioxidant property also inhibits nitrosamine formation in cured meat. The addition of ascorbic acid in flour improves baking qualities of dough and appearance of baked goods (92). Vitamin C also prevents discoloration of food during cooking and storage. Its fatty acid esters, eg, L-ascorbyl palmitate, are used to stabilize fats and oils (90). Ascorbic acid and its more stable derivatives L-ascorbyl-2-sulfate, ascorbyl polyphosphate, L-ascorbyl 2-monophosphate, etc, are added to fish feed to improve feed utilization and decrease the rate of infection (93–95). The biological and pharmacological activities of

L-ascorbyl-2-sulfate have been reviewed (96,97). Ascorbic acid is also used in agriculture as an abscission agent for fruit, in photography as a developing agent, in metallurgy as a reducing agent, in the polymer industry as a catalyst, in cosmetic formulations, in the manufacture of inks, in explosives, and in a variety of other applications (92,98).

In applications where vitamin C activity is unimportant, often D-erythorbic acid (D-araboascorbic acid) can also be used, providing the same antioxidant and reducing properties as L-ascorbic acid.

Derivatives

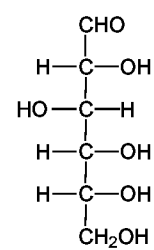
Ascorbic acid has a variety of reactive positions that can be used to synthesize derivatives (99). Various derivatives and analogues have been prepared in attempts to find substances with increased biological activity (100,101).

Only L-ascorbic acid and its salts and C-6-substituted esters have full vitamin activity; sodium L-ascorbate, calcium L-ascorbate, and L-ascorbyl palmitate are commercially significant. L-Ascorbic acid 2-sulfate is bioavailable to fish. 6-Chloro-6-deoxy-L-ascorbic acid was prepared in 1977 and its activity, compared to L-ascorbic acid, is 4–5. Derivatives, eg, 6-deoxy-L-ascorbic acid, L-ascorbic acid 3-O-methylether, and 2-amino-2-deoxy-L-ascorbic acid have been prepared; their respective activities compared to L-ascorbic acid are 1–3, 1–25–1–50, and 0. Many more derivatives with and without biological activity have been synthesized (102,103). The highest vitamin C activity correlates with the enediol lactone group, D-configuration for the C-4 hydrogen group, at least a two-carbon substituent on C-4, and L-configuration for the C-5 hydroxyl group. The primary C-6 hydroxyl group has minor impact on the biological activity.

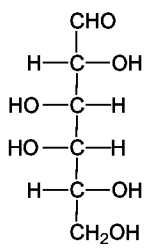
Methods for the preparation of L-ascorbic acids having isotopic C, H, O in various positions have been described and reviewed (104,105). Labeled L-ascorbic acid has played an important role in the elucidation of the metabolic pathway of L-ascorbic acid in plants and animals.

Biosynthesis

In all plants and most animals, L-ascorbic acid is produced from D-glucose (4) and D-galactose (26). Ascorbic acid biosynthesis in animals starts with D-glucose (4). In plants, where the biosynthesis is more complicated, there are two postulated biosynthetic pathways for the conversion of D-glucose or D-galactose to ascorbic acid.



(4) D-Glucose



(26) D-Galactose

Biosynthesis in Animals. Amphibians and reptiles carry the enzyme L-gulono-γ-lactone oxidase which can transform a sugar-like glucose or galactose into ascorbic acid in the kidneys. In mammals (106) and birds (107) this enzyme system has been transferred from the kidneys to the liver. Humans, other primates, guinea pigs, fruit bats, and monkeys from the top of the evolutionary tree, as well as insects and invertebrates from the bottom end of the evolutionary tree, cannot synthesize L-ascorbic acid. Thus, they must consume vitamin C from exogenous sources to survive (108). Animals that are able to produce ascorbic acid do so by the glucuronic acid pathway in the liver or kidneys. The reactions involved in rats are illustrated in Figure 8. In this pathway, the D-glucose chain remains intact, and the C-1 and C-6 of D-glucose become C-6 and C-1, respectively, of L-ascorbic acid as the sequence of carbon-chain numbering is inverted. By measuring the incorporation of ¹⁴C from D-glucose-1-¹⁴C into urinary L-ascorbic acid, it was determined that the label is found at the C-6 position of L-ascorbic acid (109).

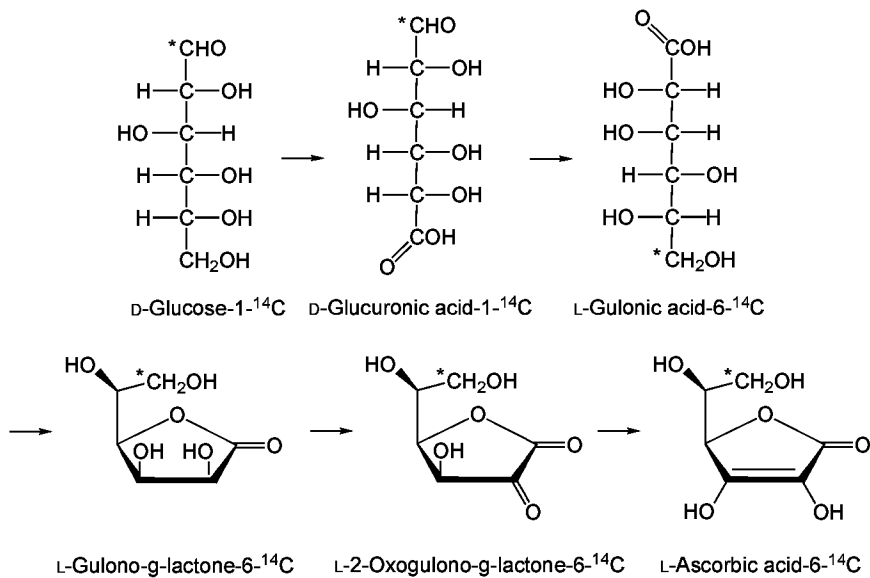


Fig. 8. Pathway for the biosynthesis of L-ascorbic acid in rats using C-1-labeled D-glucose; * indicates position of ¹⁴C.

In animals, the glucuronic pathway (Fig. 9) is an important route for glucose utilization leading to the formation of glucuronides and mucopolysaccharides. D-Glucose is first phosphorylated to D-glucose-6-phosphate, then isomerized to D-glucose-1-phosphate, which reacts with uridine-triphosphate (UTP) to form UDP-glucose. UDP-glucose is oxidized and hydrolyzed to D-glucuronic acid. D-Glucuronic acid is reduced to L-gulonic acid. L-Gulonic acid lactonizes and forms L-gulono-γ-lactone. Finally, oxidation of this intermediate is carried out by L-gulono-γ-lactone oxidase, the essential oxidizing enzyme, leading to L-ascorbic acid.

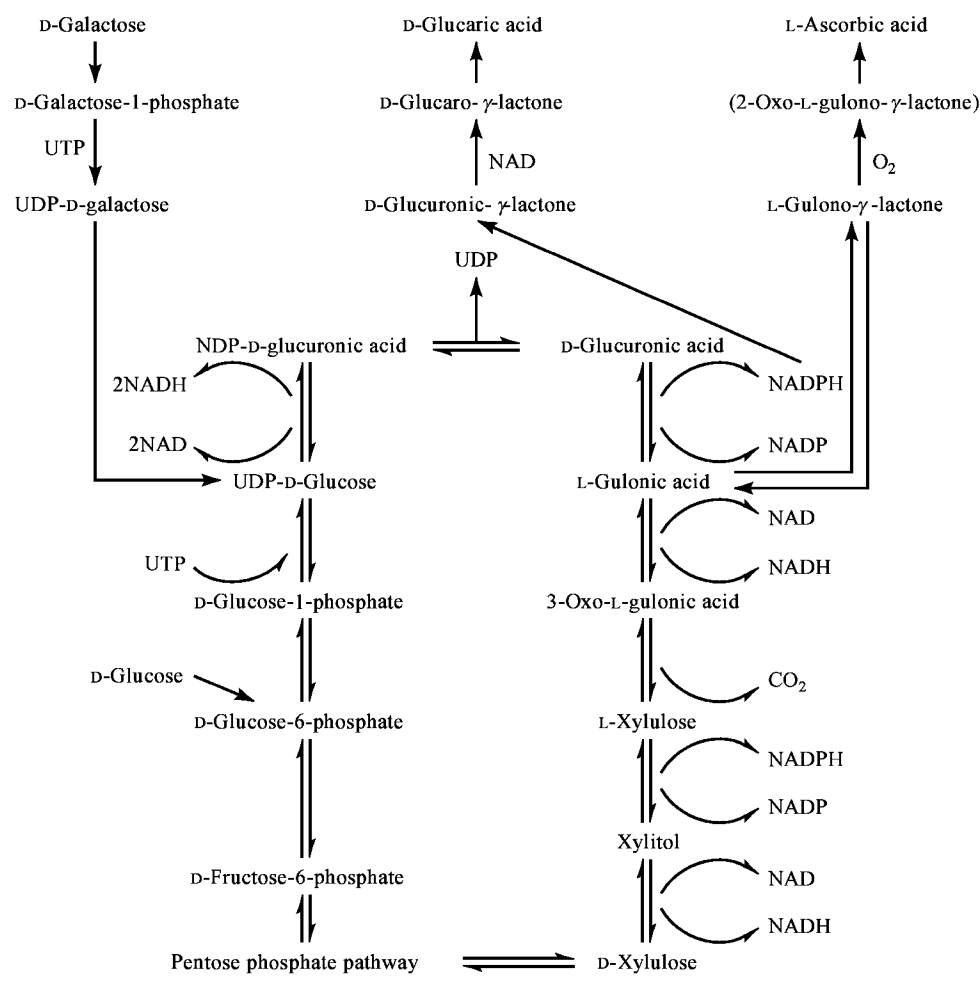


Fig. 9. Glucuronic acid pathway. NAD⁺ nicotinamide-adenine dinucleotide; NADH reduced nicotinamide-adenine dinucleotide; NADP⁺ nicotinamide-adenine dinucleotide phosphate; NADPH reduced nicotinamide-adenine dinucleotide phosphate; NDP nucleoside diphosphate; UDP uridine diphosphate; and UTP uridine triphosphate.

D-Galactose may also serve as a precursor of vitamin C because it can be converted to D-glucose. The pathway is important in the metabolism of sugars under normal and diseased conditions and in regulating physiological functions (110). The biosynthesis of L-ascorbic acid is inhibited by deficiencies of certain vitamins, eg, vitamin A, vitamin E, and biotin, but is stimulated by certain drugs, eg, barbiturates, chlorobutanol, aminopyrine, and antipyrine, and by carcinogens, eg, 3-methylcholanthrene and 3,4-benzpyrene (111–113). It has been proposed that the excretion of D-glucaric acid can be used to diagnose both the exposure to the body for foreign substances and the drug metabolic capacity of the liver (114).

Biosynthesis in Plants. As in animals, L-ascorbic acid is also the product of hexose phosphate metabolism in plants, but its biosynthesis is more complicated. There are two biosynthetic pathways for the conversion of D-glucose or D-galactose to L-ascorbic acid in plants (115). The main pathway is postulated as involving retention of configuration by oxidation at C-1 to yield D-gluconic-acid (27). Lactonization to yield D-glucono-γ-lactone (28) is followed by oxidation at C-2 or C-3 and epimerization at C-5 to afford L-2-oxogulono-γ-lactone (29) (Fig. 10). The other biosynthetic pathway from D-glucose and D-galactose may be postulated similarly to Figure 8 to account for configurational inversion (116). The main precursor of L-ascorbic acid is L-galactono-γ-lactone, rather than L-gulono-γ-lactone, which is less active (116,117), and is thought to be epimerized to L-galactono-γ-lactone prior to oxidation to L-ascorbic acid (Fig. 11). Little is known about the functions of ascorbic acid in plants. It is involved in cellular respiration and may contribute to plant growth.

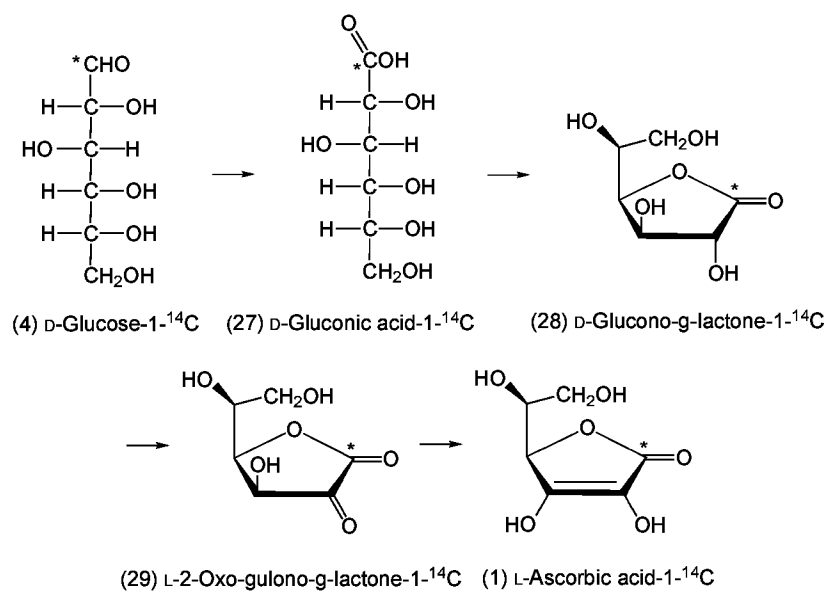


Fig. 10. Suggested pathway for the biosynthesis of L-ascorbic acid (with retention of configuration) in higher plants based on D-glucose-1-¹⁴C experiments; * indicates position of the label.

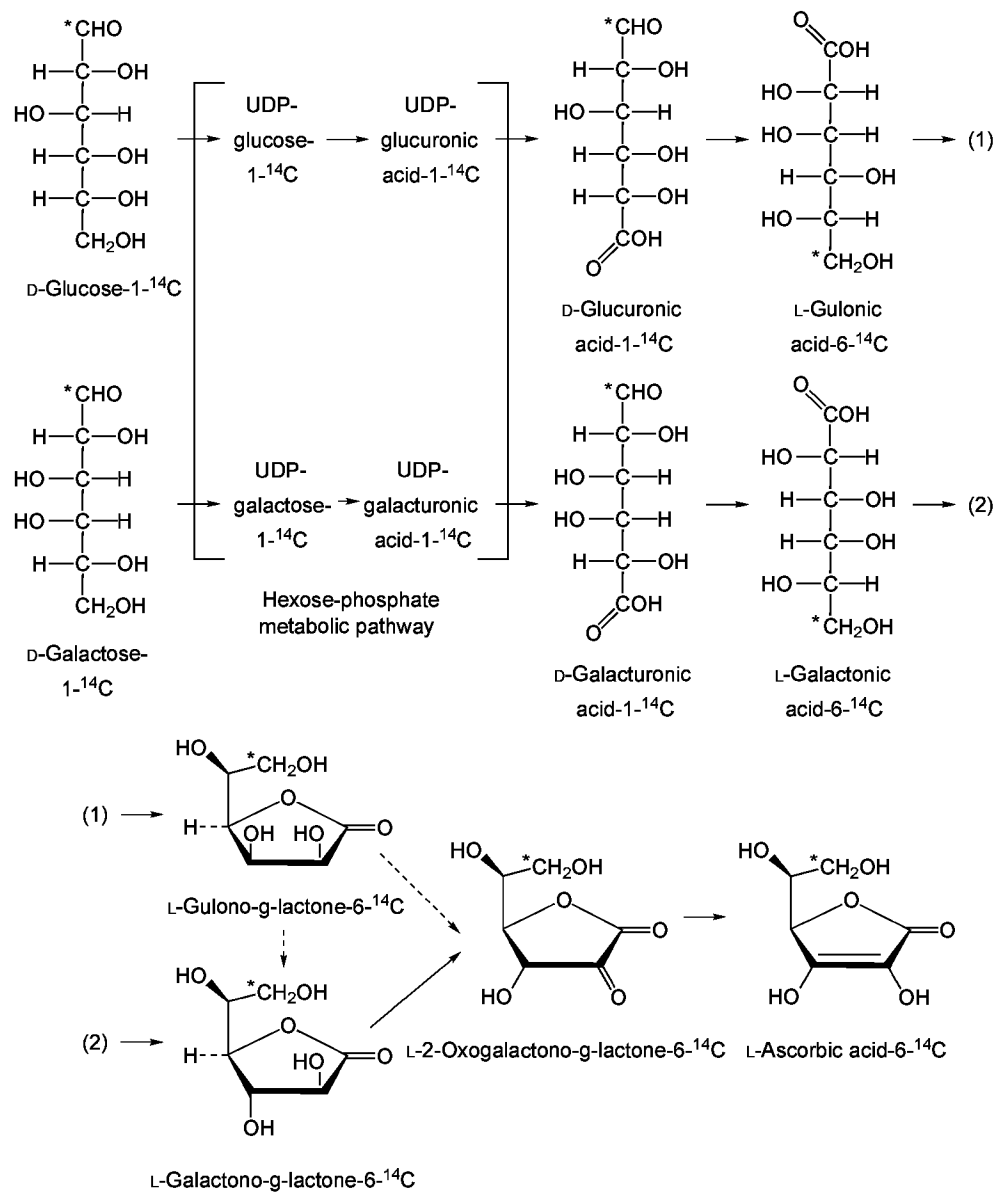


Fig. 11. Proposed pathway for the biosynthesis of L-ascorbic acid (with inversion of configuration) in plants using C-1-labeled D-glucose or C-1-labeled D-galactose; * indicates position of the ¹⁴C.

Sources of Vitamin C

Fruits and vegetables that are good sources of vitamin C include peppers, greens, broccoli, cabbage, spinach, potatoes, tomatoes, strawberries, and citrus products. Meats, fish, poultry, eggs, and dairy products contain small amounts and grain contains none (118–121). The vitamin C content of some representative foods is listed in Table 4. Potatoes and cabbage have traditionally been the most important sources of vitamin C for the majority of people in the Western World during the winter season.

Table 4. Content of L-Ascorbic Acid in Representative Foods

Food substances	L-Ascorbic acid, mg 100 g
Meat, fish, and milk	
beef, pork, fish	<2
liver, kidney	10–40
cow's milk	1–2
Vegetables	
asparagus	15–30
brussel sprouts, broccoli	90–150
cabbage	30–60
carrots	9
cauliflower	60–80
kale	120–180
leek	15–30
onion	10–30
peas, beans	10–30
parsley	170
peppers	125–200
potatoes	10–30
spinach	50–90
tomatoes	20–33
Fruit	
apples	10–30
bananas	10
grapefruit	40
guava	300
hawthorne berries	160–800
oranges, lemons	50
peaches	7–14
pineapples	17
rose hips	1000
strawberries	40–90

The contribution of any fruit and vegetable to the vitamin C content of the diet varies depending on climate, soil, and freshness. In actuality, there is a 30% coefficient of variability for ascorbic acid in fresh food products (122). In the case of field-grown spinach, for example, transportation and storage losses have been reported to be as high as 90% (122). During storage after harvesting, the vitamin C content of fruit and vegetables will decrease depending on time and temperature of storage. Ascorbic acid readily oxidizes both enzymatically and chemically on exposure to oxygen. Based on its water solubility, losses are caused through leaching during washing and blanching operations. Heat sterilization of canned foods inactivates completely the enzyme ascorbic acid oxidase which destroys vitamin C. Freezing inactivates the enzymes and is likewise a good method for preserving the vitamin C content of foodstuffs.

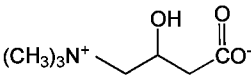
Restoration means adding to the food the amount of vitamin C which has been lost during food processing. The term *enrichment* or *fortification* is used when a nutrient is added to a food in an amount in excess of that naturally present. Fortification of foods with vitamin C is widely practiced. Technologically, it is important to avoid oxygen, copper, iron, and heat as much as possible to minimize the overages necessary above label claim to ensure compliance throughout the shelf life of the product (123). In canned foods, breakfast drinks, and cereal products (106), overages of 25–50% are commonly used.

Physiology and Biochemistry

Biochemical Functions. Ascorbic acid has various biochemical functions, involving, for example, collagen synthesis, immune function, drug metabolism, folate metabolism, cholesterol catabolism, iron metabolism, and carnitine biosynthesis. Clear-cut evidence for its biochemical role is available only with respect to collagen biosynthesis (hydroxylation of prolin and lysine). In addition, ascorbic acid can act as a reducing agent and as an effective antioxidant. Ascorbic acid also interferes with nitrosamine formation by reacting directly with nitrites, and consequently may potentially reduce cancer risk.

Enzymatic Reactions. The polypeptide collagens are the main component of skin and connective tissue, the organic substance of bones and teeth, and the ground substance between cells. The role ascorbic acid plays in collagen formation has been reviewed (124). The synthesis of collagens involves enzymatic hydroxylations of proline and of lysine. The former produces a stable extracellular matrix and the latter is needed for glycosylation and formation of cross-linkages in the fibers. Vitamin C's role in collagen formation is of importance in wound healing. Intake of 8–50 times the RDA level of 60 mg ascorbic acid per day before and after surgery increases the rate of wound healing considerably (125–127). Many of the clinical signs of scurvy are attributed to defects in collagen synthesis.

Ascorbic acid is involved in carnitine biosynthesis. Carnitine (γ-amino-β-hydroxybutyric acid, trimethylbetaine) (30) is a component of heart muscle, skeletal tissue, liver and other tissues. It is involved in the transport of fatty acids into mitochondria, where they are oxidized to provide energy for the cell and animal. It is synthesized in animals from lysine and methionine by two hydroxylases, both containing ferrous iron and L-ascorbic acid. Ascorbic acid donates electrons to the enzymes involved in the metabolism of L-tyrosine, cholesterol, and histamine (128).



(30) Carnitine

L-Tyrosine metabolism and catecholamine biosynthesis occur largely in the brain, central nervous tissue, and endocrine system, which have large pools of L-ascorbic acid (128). Catecholamine, a neurotransmitter, is the precursor in the formation of dopamine, which is converted to noradrenaline and adrenaline. The precise role of ascorbic acid has not been completely understood. Ascorbic acid has important biochemical functions with various hydroxylase enzymes in steroid, drug, and lipid metabolism. The cytochrome P-450 oxidase catalyzes the conversion of cholesterol to bile acids and the detoxification process of aromatic drugs and other xenobiotics, eg, carcinogens, pollutants, and pesticides, in the body (129). The effects of L-ascorbic acid on histamine metabolism related to scurvy and anaphylactic shock have been investigated (130). Another cellular reaction involving ascorbic acid is the conversion of folate to tetrahydrofolate. Ascorbic acid has many biochemical functions which affect the immune system of the body (131).

Antioxidant Activity. Ascorbic acid serves as an antioxidant to protect intracellular and extracellular components from free-radical damage. It

scavenges free radicals and forms the less reactive ascorbyl radical. The ascorbyl radical can be either reduced to ascorbic acid or oxidized to dehydroascorbic acid (132). Vitamin E is the major fat-soluble antioxidant involved in protecting cells from free-radical impact. During the process of fatty acid oxidation, tocopherol (vitamin E) forms the tocopheroxyl radical. Ascorbic acid has been proven to protect membrane and other hydrophobic compartments from such damage by regenerating the antioxidant form of vitamin E (131). These free-radical scavenging reactions of vitamin C are important in protecting the intracellular and extracellular structure of the lung by quenching free radicals generated by smoke, ozone, and singlet oxygen. The possibility that free-radical damage is also involved in the pathogenesis of HIV and that antioxidants may reduce infection is an area of intense interest (134).

Inhibition of Nitrosamine Formation. Nitrites can react with secondary amines and *N*-substituted amides under the acidic conditions of the stomach to form *N*-nitrosamines and *N*-nitrosamides. These compounds are collectively called *N*-nitroso compounds. There is strong circumstantial evidence that *in vivo* *N*-nitroso compounds production contributes to the etiology of cancer of the stomach (135,136), esophagus (136,137), and nasopharynx (136,138). Ascorbic acid consumption is negatively correlated with the incidence of these cancers, due to ascorbic acid inhibition of *in vivo* *N*-nitroso compound formation (139). The concentration of *N*-nitroso compounds formed in the stomach depends on the nitrate and nitrite intake. Nitrite is part of the preserving process for cured meats. Cigarette smoke contains high levels of nitrite.

Iron Absorption. A very important effect of ascorbic acid is the enhancement of absorption of nonheme iron from foods. Ascorbic acid also enhances the reduction of ferric iron to ferrous iron. This is important both in increasing iron absorption and in its function in many hydroxylation reactions (140,141). In addition, ascorbic acid is involved in iron metabolism. It serves to transfer iron to the liver and to incorporate it into ferritin. Ascorbic acid also forms soluble chelate complexes with iron (142–145). It seems ascorbic acid has no effect on high iron levels found in people with iron overload (146). It is well known, in fact, that ascorbic acid in the presence of iron can exhibit either prooxidant or antioxidant effects, depending on the concentration used (147). The combination of citric acid and ascorbic acid may enhance the iron load in aging populations. Iron overload may be the most important common etiologic factor in the development of heart disease, cancer, diabetes, osteoporosis, arthritis, and possibly other disorders. The synergistic combination of citric acid and ascorbic acid needs further study, particularly because the iron overload produced may be correctable (147).

Deficiency. Scurvy is a vitamin C-specific disease. It is characterized by anemia and alteration of protein metabolism; weakening of collagenous structures in bone, teeth, and connective tissues; swollen, bleeding gums with loss of teeth; fatigue and lethargy; rheumatic pains in the legs and degeneration of the muscles, skin lesions, and capillary weakness, massive hematomas in the thighs; and hemorrhages in many organs, including the eyes. Small (10–60 mg d) quantities of L-ascorbic acid are sufficient to reverse the trend of both subclinical and clinical scurvy and alleviate their symptoms. Plasma and leukocyte ascorbic acid levels are the most reliable markers of vitamin C intake. Leukocyte levels are more reliable, less sensitive to recent vitamin C intake, and better reflect tissue ascorbate, but require relatively large amounts of blood. Normal leukocyte vitamin C levels are 20–40 µg 10⁸ cells (148). Plasma concentrations of ascorbic acid less than 0.2 mg dL and leukocyte concentrations below 2 µg 10⁸ cells are seen in scurvy (149). Plasma ascorbic acid levels of 0.5 mg dL are considered to prevent deficiency symptoms. Normal plasma levels are 0.8–1.4 mg dL.

Absorption, Transport, and Excretion. The vitamin is absorbed through the mouth, the stomach, and predominantly through the distal portion of the small intestine, and hence, penetrates into the bloodstream. Ascorbic acid is widely distributed to the cells of the body and is mainly present in the white blood cells (leukocytes). The ascorbic acid concentration in these cells is about 150 times its concentration in the plasma (150,151). Dehydroascorbic acid is the main form in the red blood cells (erythrocytes). White blood cells are involved in the destruction of bacteria.

Up to 80% of oral doses of ascorbic acid are absorbed in humans with intakes of less than 0.2 g of vitamin C. Absorption of pharmacological doses ranging from 0.2 g to 12 g results in an inverse relationship, with less than 20% absorption at the higher doses. A single oral dose of 3 g has been reported to approach the absorptive capacity (tissue saturation) of the human intestine. Higher blood levels can be attained by providing multiple divided vitamin C doses per day.

The adrenal glands and pituitary glands have the highest tissue concentration of ascorbic acid. The brain, liver, and spleen, however, represent the largest contribution to the body pool. Plasma and leukocyte ascorbic acid levels decrease with increasing age (152). Elderly people require higher ascorbic acid intakes than children to reach the same plasma and tissue concentration (153).

Ascorbic acid is very soluble in water and mainly excreted in the urine. No ascorbic acid is excreted during vitamin C deficiency. A minimum amount is lost in the feces, even after intake of gram dosages (154).

Mobilization and Metabolism. The total ascorbic acid body pool in healthy adults has been estimated to be approximately 1.5 g, which increases to 2.3–2.8 g with intakes of 200 mg d (151–158). Depletion of the body pool to 600 mg initiates physiological changes, and signs of clinical scurvy are reported when the body pool falls below 300 mg (149). Approximately 3–4% of the body pool turns over daily, representing 40–60 mg d of metabolized, or consumed, vitamin C. Smokers have a higher metabolic turnover rate of vitamin C (approximately 100 mg d) and a lower body pool than nonsmokers, unless compensated through increased daily intakes of vitamin C (159). The metabolism of ascorbic acid varies among different species.

In rats and guinea pigs, respiratory carbon dioxide is the major oxidation product of vitamin C (160). Given to humans in physiological doses, approximately 10% of a dose will be recovered in the urine, and urinary oxalic acid is the predominant metabolite. The formation of urinary metabolites is limited to 30–50 mg d. Intake of up to 10 g d ascorbic acid does not result in a considerable increase of urinary metabolites. The vitamin is excreted largely unmetabolized (161). In humans, the metabolites are dehydroascorbic acid, oxalate, 2,3-diketo-L-gulonic acid and derivatives of L-threose and L-threonic acid. L-Ascorbate-2-sulfate has been found in human urine as well (162). The feces are not a significant excretory route unless doses over 1 g are given.

The half-life of ascorbic acid is inversely related to the daily intake and is 13–40 d in humans and 3 d in guinea pigs, which is consistent with the longer time for humans to develop scurvy.

Requirements. The level of ascorbic acid intake required is dependent in part upon the body's handling of the nutrient. The recommendations for the daily intake of vitamin C in various countries range from 30 to 100 mg d. There is an extensive lack of knowledge about the biochemical and physiological functions of vitamin C. Although as little as 10 mg d of ascorbic acid can prevent clinical scurvy, this intake is insufficient to maintain an adequate body pool of the vitamin for peak physical and mental health. The RDA (Recommended Dietary Allowances) for vitamin C in the United States is 60 mg for men and women to maintain the body pool (Table 5). Vitamin C levels are higher for pregnant and lactating women to account for losses to the fetus and to breast milk.

Table 5. The 1989 RDA and US RDA for Vitamin C

Group	Age, yr	Vitamin C, mg
	RDA	
infants	0–0.5	30
	0.5–1.0	35
children	1–3	40
	4–10	45
males	11–14	50
	15–51 +	60
females	11–14	50
	15–51 +	60
pregnant		70
lactating		
1st 6 mo		95
2nd 6 mo		90
cigarette smokers		100
	U.S. RDA	
infants under 13 mo		35

children under 4 yr	40
adults and children over 4 yr	60
pregnant or lactating women	60

The most recent RDA has included a vitamin C recommendation of 100 mg day for cigarette smokers. An increasing number of investigators have concluded that the current RDA for vitamin C may not be adequate for elderly individuals. Plasma vitamin C level is generally accepted as an indicator of vitamin C status.

A recent study recommended that the current RDA be increased from 60 mg d to 200 mg. The researchers indicated, however, that vitamin C daily doses above 400 mg have no value (163).

Toxicity. The acceptable daily allowance, which may be ingested without any risk of harm, is 1050 mg for a 70-kg healthy person (20). There is also no evidence in the literature that ingestion of up to 10 g vitamin C per day constitutes a serious health risk for humans.

BIBLIOGRAPHY

"Ascorbic Acid" in *ECT* 1st ed., Vol. 2, pp. 150–163, by P. Regna, Chas. Pfizer & Co., Inc.; in *ECT* 2nd ed., Vol. 2, pp. 747–762, by J. J. Burns, Burroughs Wellcome & Co. (U.S.A.), Inc.; "Ascorbic Acid" under "Vitamins" in *ECT* 3rd ed., Vol. 24, pp. 8–40, by G. M. Jaffe, Hoffmann-La Roche Inc.

1. *Biochem. Biophys. Acta* **107**, 4 (1965).
2. W. N. Haworth, *Chem. Ind.* (London) **52**, 482 (1993); W. N. Haworth and E. L. Hirst, *Ergeb. Vitamin-u: Hormonforsch.* **2**, 160 (1939).
3. R. E. Hodges and E. M. Baker in R. S. Goodhart and M. E. Shils, eds., *Modern Nutrition in Health and Disease*, 5th ed., Lea & Febiger, Philadelphia, Pa., 1973, p. 245; U. Wintermeyer, H. Lahann, and R. Vogel, *Vitamin C: Entdeckung, Identifizierung und Synthese—Heutige Bedeutung in Medizin und Lebensmitteltechnologie*, Deutscher Apotheker Verlag, Stuttgart, Germany, 1981.
4. L. Stone, *The Healing Factor—Vitamin C Against Disease*, Grosset & Dunlap, Inc., New York, 1972, p. 7.
5. I. M. Sharman in G. G. Birch and K. J. Parker, eds., *Vitamin C: Recent Aspects of its Physiological and Technological Importance*, John Wiley & Sons, Inc., New York, 1974, p. 1.
6. A. F. Hess, *Scurvy, Past and Present*, J. B. Lippincott Co., Philadelphia, Pa., 1920.
7. J. Lind, *A Treatise on the Scurvy*, London, 1753 (republished by Edinburgh University Press, Edinburgh, Scotland, 1953).
8. A. Holst and T. Fröhlich, *J. Hyg.* **7**, 634 (1907).
9. H. Chick and E. M. Hume, *J. R. Army Med. Corps* **29**, 121 (1917); *J. Biol. Chem.* **39**, 203 (1919).
10. A. Harden and S. S. Zilva, *Biochem. J.* **12**, 259 (1918); S. S. Zilva and F. M. Wells, *Proc. R. Soc. London Ser. B* **90**, 505 (1917–1919); S. S. Zilva, *Biochem. J.* **21**, 689 (1927).
11. A. Szent-Györgyi, *Biochem. J.* **22**, 1387 (1928); J. L. Svribely and A. Szent-Györgyi, *Biochem. J.* **26**, 865 (1932).
12. C. G. King and W. A. Waugh, *Science* **75**, 357 (1932).
13. W. N. Haworth and E. L. Hirst, *J. Soc. Chem. Ind.* **52**, 645–646 (1933).
14. T. Reichstein, A. Grössner, and R. Oppenhauer, *Helv. Chim. Acta* **16**, 1019–1033 (1933).
15. T. Reichstein and A. Grössner, *Helv. Chim. Acta* **17**, 311 (1934).
16. P. Karrer, K. Schnupp, and F. Zehner, *Helv. Chim. Acta* **16**, 1161–1163 (1933).
17. F. Michael and K. Kraft, Hoppe-Seyler's *Z. Physiol. Chem.* **222**, 235–249 (1933).
18. H. von Euler and C. Martius, *Ann.* **505**, 73–87 (1933).
19. E. L. Hirst, *Fortsch. Chem. Org. Naturst.* **2**, 132–159 (1939).
20. *Wld. Hlth. Org. Techn. Rep. Ser.* 1972, No. 505; *Wld. Hlth. Org. Techn. Rep. Ser.* 1974, No. 539.
21. Armour Research Foundation, *Anal. Chem.* **20**, 986–987 (1948).
22. C. von Planta, *Helv. Chim. Acta* **44**, 1444–1446 (1961).
23. J. Hvoslef, *Acta Crystallogr. Sect. B* **24**, 1431–1440 (1968).
24. J. Hvoslef, *Acta Chem. Scand.* **18**, 841–842 (1964).
25. J. Hvoslef, *Acta Crystallogr. Sect. B* **24**, 23–35 (1968).
26. W. Müller-Mulot, Hoppe-Seyler's *Z. Physiol. Chem.* **351**, 52,56 (1970); J. Hvoslef, *Acta Chem. Scand.* **24**, 2238 (1970); *Acta Crystallogr.* **B28**, 916 (1972); J. Hvoslef and B. Pedersen, *Acta Chem. Scand.* **B33**, 503 (1979); *Carbohydr. Res.* **92**, 9 (1981).
27. R. J. Kutsky, *Handbook of Vitamins and Hormones*, Van Nostrand Reinhold Co., Inc., New York, 1973, p. 71; *The Merck Index*, 11th ed., Merck & Co., Inc., Rahway, N.J., 1989, p. 130.
28. J. S. Koevendel, *Nature* **180**, 434 (1957); Y. Ogata and Y. Kosugi, *Tetrahedron* **26**, 4711 (1970).
29. Sadtler Standard Spectra, *Spectrum No. 5424*, Sadtler Research Laboratory, Inc., Philadelphia, Pa., 1972.
30. High Resolution NMR Spectra Catalog, Vol. 2, *Spectrum No. 464*, Varian Assoc., Palo Alto, Calif., 1963.
31. V. F. Johnson and W. C. Jankowski, Carbon-13 NMR Spectra, *Spectrum No. 171*, Wiley-Interscience, New York, 1972; S. Berger, *Tetrahedron* **33**, 1587 (1977); T. Ogama, J. Uzawa, and M. Matsui, *Carbohydr. Res.* **59**, C32 (1977); T. Radford, J. G. Sweeny, and G. A. Jacobucci, *J. Org. Chem.* **44**, 658 (1979).
32. B. M. Tolbert and co-workers, *Ann. N.Y. Acad. Sci.* **258**, 44 (1975); B. M. Tolbert, *Int. J. Vit. Nutr. Res. Suppl.* **19**, 127 (1979).
33. B. Tolbert and J. B. Ward, in *Ascorbic Acid, Chemistry, Metabolism and Uses*, *Adv. Chem. Sci.* **200**, 101–123 (1982); M. Ohmori, M. Takage, *Agr. Biol. Chem.* **42**, 173–174 (1978).
34. F. E. Huelin, *Food Res.* **18**, 633 (1953).
35. J. Campbell and W. G. Tubb, *Can. J. Res.* **28E**, 19–32 (1950).
36. K. Mikova and J. Davidele, *Chem. Listy* **68**, 715 (1974).
37. F. E. Huelin, *Food Res.* **18**, 633 (1953); J. Campbell and W. G. Tubb, *Can. J. Res.* **28E**, 19 (1947).
38. D. W. Bradley, G. Emery, and J. E. Maynard, *Clin. Chim. Acta* **4**, 47–52 (1973).
39. T. Reichstein, A. Grössner, and R. Oppenauer, *Helv. Chim. Acta* **16**, 561–565 (1933).
40. T. Reichstein, A. Grössner, and R. Oppenauer, *Nature (London)* **132**, 280 (1932).
41. T. Reichstein, A. Grössner, and R. Oppenauer, *Helv. Chim. Acta* **17**, 510–520 (1934).
42. **U.S. Pat. 2,056,126** (Sept. 29, 1936), T. Reichstein (to Hoffmann-La Roche Inc.).
43. W. N. Haworth, *Chem. Ind. (London)* **52**, 482–485 (1933).
44. R. G. Ault and co-workers, *J. Chem. Soc.* 1419–1423 (1933).
45. W. N. Haworth and E. L. Hirst, *Helv. Chim. Acta* **17**, 520–523 (1934).
46. T. C. Crawford and S. A. Crawford, *Adv. Carbohydr. Chem. Biochem.* **37**, 79 (1980).
47. M. Kulhanek, *Adv. Appl. Microbiol.* **12**, 11–33 (1970).
48. G. Xin, *Weishengwuxue Tongbao* **14**(2), 73 (1987).
49. **Eur. Pat. Appl. 278 447** (Prior. Feb. 2, 1987) W. Ning, Z. Tao, C. Wang, S. Wang, Z. Yan, and G. Yin (to Institute of Microbiology Academia Sinica).

50. **Eur. Pat. Appl. 221 707** (Prior. Oct. 22, 1985) I. Nogami, H. Shirafuji, M. Oka, and T. Yamaguchi (to Takeda Chemical Industries, Ltd.).
51. T. Sonoyama and co-workers, *Appl. Environ. Microbiol.* **43**, 1064 (1982).
52. S. Anderson and co-workers, *Science* **230**, 144–149 (1985).
53. J. F. Grindley, M. A. Payton, H. van do Pol, and K. G. Hardy, *Appl. Environ. Microbiol.* **54**, 1770 (1988).
54. **Eur. Pat. Appl. No. 207 763** (1985), T. J. Skatrud and R. J. Huss (to Bio-Technical Resources).
55. I. T. Strukov and N. A. Kopylova, *Farmatsiya* **10**(3), 8 (1974); *Drug Cosmet. Ind.* 40 (Aug. 1977).
56. W. Ipatiew, *Ber.* **45**, 3218 (1912).
57. L. W. Cover, R. Connor, and H. Adkins, *J. Am. Chem. Soc.* **54**, 1651 (1932).
58. A. T. Kuhn, *J. Appl. Electrochem.* **11**, 261 (1981).
59. W. E. Cake, *J. Am. Chem. Soc.* **44**, 859 (1922).
60. M. Kulhanek, *Adv. Appl. Microbiol.* **12**, 11 (1970).
61. P. A. Wells, J. J. Stubbs, L. B. Lockwood, and E. T. Roe, *Ind. Eng. Chem.* **29**, 1385 (1937); E. J. Fulmer and L. A. Underkofler, *Iowa State Coll. J. Sci.* **21**, 251 (1974).
62. **U.S. Pat. 3,622,560** (Nov. 23, 1971), N. Halder, M. J. O'Leary, and N. C. Hindley; **U.S. Pat. 3,598,804** (Aug. 10, 1971), N. Halder, N. C. Hindley, G. M. Jaffe, M. J. O'Leary, and P. Weinert; **U.S. Pat. 3,607,862** (Sept. 21, 1971), G. M. Jaffe, W. Szkrybalo, and P. Weinert (all to Hoffmann-La Roche Inc.).
63. R. S. Glass, S. Kwoh, and E. P. Oliveto, *Carbohydr. Res.* **26**, 181 (1973).
64. **U.S. Pat. 2,039,929** (May 5, 1936), T. Reichstein, **U.S. Pat. 2,301,811** (Nov. 10, 1942), T. Reichstein (both to Hoffmann-La Roche Inc.); V. F. Kazimirova, *J. Gen. Chem. (USSR)* **24**, 1559 (1955).
65. J. W. Weijlard, *J. Am. Chem. Soc.* **67**, 1031 (1945); **U.S. Pat. 2,367,251** (Jan. 16, 1945), J. Weijlard and J. B. Ziegler (to Merck and Co.); K. Nakagawa, R. Konaka, and T. Nakata, *J. Org. Chem.* **27**, 1597 (1962).
66. L. O. Shnaidman and I. N. Kushehinskaya, *Tr. Vses. Nauchno-Issled. Vitam. Inst.* **8**, 13 (1961).
67. **U.S. Pat. 3,832,355** (Aug. 27, 1972), G. M. Jaffe and E. Plevén (to Hoffmann-La Roche Inc.).
68. **U.S. Pat. 3,453,191** (July 1, 1969), G. J. Fröhlich, A. J. Kratavil, and E. Zrike (to Hoffmann-La Roche Inc.); **Eur. Pats. 40331** (Nov. 25, 1981) and **40709** (Dec. 2, 1981), R. Whittmann, W. Wintermeyer, and J. Butska (to Merck GMBH); P. M. Robertson, P. Berg, H. Reimann, K. Schleich, and P. Seiler, *Chimia* **36**, 305 (1982).
69. P. P. Regna and B. P. Caldwell, *J. Am. Chem. Soc.* **66**, 246 (1944).
70. F. Elger, *Festschr. Emil Barends*, 229 (1936); **Brit. Pat. 428,815** (May 20, 1935); and **Brit. Pat. 466,548** (May 31, 1937), T. Reichstein (to Hoffmann-La Roche, Inc.).
71. *The United States Pharmacopeia/National Formulary (USP XXIII/NFXVIII)*, United States Pharmacopeial Convention, Inc., Rockville, Md., 1995, pp. 130, 245, 1414, 2215.
72. Food and Nutrition Board, National Research Council, *Food Chemicals Codex*, 3rd ed., National Academy Press, Washington, D.C., 1981, pp. 27, 28, 277–278.
73. L. A. Pachla and D. L. Reynolds, *J. Assoc. Off. Anal. Chem.* **68**, 1–12 (1985).
74. J. R. Cooke and R. E. D. Moxon, in *Vitamin C (Ascorbic Acid)*, J. N. Counsell and D. H. Hornig, eds., Applied Science Publishers, London, 1981, pp. 167–198; *Adv. Chem. Ser.* **200**, 199 (1982); R. C. Rose and D. L. Nahrwold, *Anal. Biochem.* **123**, 389 (1982).
75. C. I. Bliss and P. György, in P. György, ed., *Vitamin Methods*, Vol. II, Academic Press, New York, 1951, pp. 244–261.
76. J. Tillmans, *Z. Unters. Lebensmittel* **54**, 33 (1927).
77. R. Strohecker, *Wiss. Veröff. Dtsch. Ges. Ernähr.* **14**, 157 (1965).
78. T. Tono and S. Fujita, *Agric. Biol. Chem.* **45**, 2947 (1981).
79. K. Dabrowski and co-workers, *J. Aquaculture* **124**, 169–192 (1994).
80. J. H. Roe and C. A. Kuether, *J. Biol. Chem.* **147**, 399–407 (1943).
81. M. J. Deutsch and C. E. Weeks, *J. Assoc. Off. Anal. Chem.* **48**, 1248–1256 (1965).
82. G. Brubacher, W. Möller-Mulot, and D. A. T. Southgate, *Methods for the Determination of Vitamins in Food Recommended by Cost 91*, Elsevier Applied Science, New York, 1985, Chapt. 7.
83. J. Geigert, D. S. Hirano, and S. L. Neidleman, *J. Chromatogr.* **206**, 396–399 (1981).
84. S. Garcia-Castineiras, V. D. Bonnet, R. Figueroa, and M. Miranda, *J. Liq. Chromatogr.* **4**, 1619–1640 (1981).
85. W. Lee, P. Hamernyik, M. Hutchinson, V. A. Raisys, and R. F. Labbe, *Clin. Chem.* **28**, 2165–2169 (1982).
86. H. J. Kim, *J. Assoc. Official Anal. Chem.* **72**, 681–686 (1989).
87. L. L. Lloyd, F. Watner, J. F. Kennedy, and C. A. White, *J. Chromatogr.* **437**, 453–456 (1988).
88. O. Pelletier in J. Augustin, B. P. Klein, D. A. Becker, and P. B. Venugopal, eds., *Methods of Vitamin Assay*, 4th ed., John Wiley & Sons, Inc., New York, 1985, Chapt. 12.
89. M. H. Bui-Nguyen, in A. P. DeLeenheer, eds., *Modern Chromatographic Analysis of the Vitamins*, Marcel Dekker, New York, 1985, Chapt. 5.
90. B. Borenstein, *Food Tech.* **41**, 98–99 (1987).
91. M. Liao and P. A. Seib, *Food Tech.* **41**, 104–107 (1987).
92. J. C. Bauernfeind, in *Ascorbic Acid: Chemistry, Metabolism, and Uses*, P. A. Seib and B. M. Tolbert, eds., American Chemical Society, Washington, D.C., 1982, pp. 395–497.
93. V. S. Durve and R. T. Lovell, *Can. J. Fish Aquat. Sci.* **39**, 948–951 (1982).
94. M. Matusiewicz, K. Dabrowski, L. Volker, and K. Matusiewicz, *J. Nutr.* **125**, 3055–3061 (1995).
95. R. T. Lovell and G. O. Elnaggar, *Proceedings of the Third International Symposium on Feeding and Nutrition in Fish*, Toba, Japan, (Aug. 28–Sept. 1, 1989), pp. 159–165.
96. G.-O. Toa, *Sheng Lik'o Hseuh Chin Chan* **12**, 26 (1981).
97. K. Dabrowski, M. Matusiewicz, and J. H. Blom, *Aquaculture* **124**, 169–190 (1994).
98. **U.S. Pat. No. 4,728,376** (Mar. 1, 1988), E. F. Kurtz (to Golden Powder of Texas, Inc.).
99. O. Isler, G. Brubacher, S. Ghisla, and B. Kaudler, *Vitamin H*, Georg Thieme Verlag, Stuttgart, Germany, 1988, pp. 396–444.
100. F. Smith, *Adv. Carbohydr. Chem. Biochem.* **2**, 79 (1946).
101. J. Stanek, M. Cerny, J. Kocourek, and J. Pacak, *The Monosaccharides*, Academic Press, Inc., New York, 1963, p. 735.
102. G. Mohn, in W. Dirscherl and W. Friedrich, eds., *Fermente, Hormone, Vitamine, und die Beziehungen dieser Wirkstoffe zueinander*, 3rd ed., Vol. 3, Pt. 1 (Vitamine), Thieme, Stuttgart, Germany, 1982, pp. 910–913.
103. R. S. Harris, *The Vitamins*, 2nd ed., Vol. 1, Academic Press, New York, London, 1967, 383–385.
104. M. Williams and F. A. Loewus, *Carbohydr. Res.* **63**, 149 (1978).
105. S. L. von Schuching and A. F. Abt, *Methods Enzymol.* **18**, 1 (1970).
106. I. B. Chatterjee, *Science* **82**, 1271–1272 (1973).
107. I. B. Chatterjee, A. K. Majumder, B. K. Nandi, and N. Subramanian, in *Second Conference on Vitamin C*, Vol. 258, C. G. King and J. J. Burns, eds., New York Academy of Sciences, New York, 1975, pp. 24–47.
108. J. J. Burns, *Am. J. Med.* **26**, 740 (1959).
109. F. A. Isherwood, Y. T. Chen, and L. W. Mapson, *Biochem. J.* **56**, 11 (1954); J. J. Burns and C. Evans, *J. Biol. Chem.* **23**, 897 (1956).

110.

G. J. Dutton, ed., *Glucuronic Acid–Free and Combined: Chemistry, Biochemistry, Pharmacology and Medicine*, Academic Press, Inc., New York 1966; B. L. Horecker in G. Ritzel and G. Brubacher, eds., *Monosaccharides and Polyols in Nutrition, Therapy and Dietetics*, Hans Huber, Bern, Switzerland, 1967, p. 1.

111.

A. H. Conney, G. A. Bray, C. E. Evans, and J. J. Burns, *Ann. N.Y. Acad. Sci.* **92**, 115 (1961).

112.

J. J. Burns, C. Evans, and N. Trousof, *J. Biol. Chem.* **227**, 785–794 (1957).

113.

F. Horio and A. Yoshida, *J. Nutr.* **112**, 416–425 (1982).

114.

W. R. F. Notten, Stichting Studentenpers, University of Nijmegen, Nijmegen, The Netherlands, 1975; W. R. F. Notten and P. T. Henderson, *Int. Arch. Occup. Environ. Health* **38**, 197, 209 (1977).

115.

F. A. Loewus and M. M. Baig, *Methods Enzymol.* **18**, 22 (1970); F. A. Loewus, G. Wagner, and J. C. Yang, *Ann. N.Y. Acad. Sci.* **258**, 7 (1975); F. A. Loewus in Preiss, ed., *Biochemistry of Plants*, Vol. 3, Academic Press, Inc., New York, 1980, p. 77; C. H. Foyer and B. Halliwell, *Phytochemistry* **16**, 1347 (1977).

116.

F. A. Isherwood, Y. T. Chen, and L. W. Mapson, *Biochem. J.* **56**, 1 (1954).

117.

M. M. Baig, S. Kelley, and F. A. Loewus, *Plant. Physiol.* **46**, 277 (1970).

118.

Committee on Dietary Allowances, Food and Nutrition Board, *Recommended Dietary Allowances*, 10th ed., National Academy of Sciences, Washington, D.C., 1989, pp. 115–124.

119.

W. H. Sebrell and R. S. Harris, eds., *The Vitamins, Chemistry, Physiology, Pathology, Methods*, 2nd ed., Vol. 1, Academic Press, Inc., New York, 1967, p. 305.

120.

J. Marks, *A Guide to the Vitamins, Their Role in Health and Disease*, Medical and Technical Publishing Co., Lancaster, U.K., 1975.

121.

S. Nobile and J. M. Woodhill, *Vitamin C: The Mysterious Redox-System—A Trigger of Life*, MTP Press, Boston, Mass., 1981, p. 38.

122.

P. A. La Chance and M. C. Fisher, in R. Karmas and R. S. Harris, eds., *Nutritional Evaluation of Food Processing*, 3rd ed., Van Nostrand Publishing Co., New York, 1988, p. 506.

123.

B. Borenstein, in T. E. Furis, ed., *Handbook of Food Additives*, 2nd ed., CRC Press, Cleveland, Ohio, 1972, pp. 91, 101.

124.

B. Peterkofsky and S. Udenfriend, *Proc. Nat. Acad. Sci.* **53**, 335 (1965); M. J. Barnes and E. Kodicek, *Vitam. Horm.* **30**, 1 (1972).

125.

A. H. Hunt, *Br. J. Surg.* **28**, 436–461 (1941).

126.

S. P. Shukla, *Experientia* **25**, 704 (1969).

127.

T. T. Irvin, D. K. Chattopadhyay, and A. Smythe, *Surg. Gynec. Obstet.* **147**, 49–55 (1978).

128.

C. J. Bates, in J. N. Counsell and D. H. Hornig, eds., *Vitamin C (Ascorbic Acid)*, Applied Science Publishers, London, 1981.

129.

V. G. Zannoni, J. I. Brodfuehrer, R. C. Smart, and R. L. Susick, in J. J. Burns, J. M. Rivers, and L. J. Machlin, eds., *Third Conference on Vitamin C*, N.Y. Academy of Sciences, New York, 1987.

130.

W. Dawson and G. B. West, *Br. J. Pharmacol.* **24**, 75 (1965).

131.

A. Bendich, *Food Tech.* **41**, 112–114 (1987).

132.

A. Bendich, L. J. Machlin, O. Scandurra, G. W. Burton, and D. D. M. Wayner, *Adv. in Free Radical Biol. Med.* **2**, 419–444 (1986).

133.

R. E. Beyer, *J. Bioenerget. Biomembr.*, **26**(4), 349–358 1994.

134.

R. J. Jariwalla, S. Harakch, in R. R. Watson, ed., *Nutrition and AIDS*, CRC Press, Boca Raton, Fla., 1994, pp. 117–139.

135.

S. S. Mirvish, *J. Natl. Cancer Inst.* **71**, 629–647 (1983).

136.

P. N. Magee, *Cancer Surv.* **8**, 207–239 (1989).

137.

V. M. Craddock, *Eur. J. Cancer Prev.* **1**, 89–103 (1992).

138.

Y. Zeng and co-workers, in *Cancer Epidemiol. Biomarkers & Prev.* **2**, 195–200 (1993).

139.

S. S. Mirvish, *Cancer Res. (Suppl.)*, 1948–1951 (Apr. 1, 1994).

140.

L. Hallberg, in *Vitamin C (Ascorbic Acid)*, J. N. Counsell and D. H. Hornig, eds., Applied Science Publishers, London, 1981, pp. 49–61.

141.

L. Hallberg, M. Brune, and L. Rossander-Hulthen, in J. J. Burns, J. M. Rivers, and L. J. Machlin, eds., *Third Conference on Vitamin C*, Vol. 498, New York Academy of Sciences, New York, 1987, pp. 324–332.

142.

E. R. Monsen and J. F. Page, *J. Agric. Food Chem.* **26**, 223–226 (1978).

143.

N. W. Solomons and F. E. Viteri, in P. A. Seib and B. M. Tolbert, eds., *Ascorbic Acid: Chemistry, Metabolism, and Uses*, American Chemical Society, Washington, D.C., 1982, pp. 551–569.

144.

C. M. Plug, D. Decker, and A. Bult, *Pharm. Weekblad* **16**, 245–248 (1984).

145.

J. J. M. Marx and J. Stiekema, *Eur. J. Clin. Pharmacol.* **23**, 335–338 (1982).

146.

A. Bendich in L. Packer and J. Fuchs, eds., *Vitamin C in Health and Disease*, J. Marcel Dekker Inc., NY, 1997, pp. 411–416.

147.

R. D. Crawford, *Biochem. Molec. Med.* **54**, 1–11 (1995).

148.

C. J. Schorah, in T. G. Tayler, ed., *The Importance of Vitamins to Health*, MTP Press, Lancaster, U.K., 1979, pp. 61–72.

149.

R. E. Hodges, E. M. Baker, J. Hood, H. E. Sauberlich, and S. C. March, *Am. J. Clin. Nutr.* **22**, 535–548 (1969).

150.

R. Anderson, in Ref. 128, pp. 249–272.

151.

A. Castelli, G. E. Martorana, E. Meucci, and G. Bonetti, *Acta Vitaminol. Enzymol.* **4**, 189–196 (1982).

152.

C. J. Schorah and co-workers, *Am. J. Clin. Nutr.* **34**, 871–876 (1981).

153.

P. J. Garry, J. S. Goodwin, W. C. Hunt, and B. A. Gilbert, *Am. J. Clin. Nutr.* **36**, 332–339 (1982).

154.

D. Hornig, J. P. Vuileumier, and D. Hartmann, *Intern. J. Vitamin Nutr. Res.* **50**, 309–314 (1980).

155.

E. M. Baker, J. C. Saari, and B. M. Tolbert, *Am. J. Clin. Nutr.* **19**, 371–378 (1966).

156.

E. M. Baker, R. E. Hodges, J. Hood, H. E. Sauberlich, and S. C. March, *Am. J. Clin. Nutr.* **22**, 549–558 (1969).

157.

E. M. Baker, R. E. Hodges, J. Hood, H. E. Sauberlich, and E. M. Baker, *Am. J. Clin. Nutr.* **24**, 444–454 (1971).

158.

R. E. Hodges, J. Hood, J. Canham, H. E. Sauberlich, and E. M. Baker, *Am. J. Clin. Nutr.* **24**, 432–443 (1971).

159.

A. B. Kallner, D. Hartmann, and D. Hornig, *Am. J. Clin. Nutr.* **34**, 1347–1355 (1981).

160.

D. Hornig and D. Hartmann, in P. A. Seib and P. M. Tolbert, eds., *Ascorbic Acid: Chemistry, Metabolism and Uses*, American Chemical Society, Washington, D.C., 1982, pp. 293–316.

161.

A. Kallner, D. Hartmann, and D. Hornig, *Am. J. Clin. Nutr.* **32**, 530–539 (1979).

162.

E. M. Baker, D. C. Hammer, S. C. March, B. M. Tolbert, and J. E. Canham, *Science* **173**, 826–827 (1971).

163.

M. Levine and co-workers, *Proc. Natl. Acad. Sci. USA* **93**, 3704 (1996).

General References

Committee on Dietary Allowances, Food and Nutrition Board, *Recommended Dietary Allowances*, 10th ed., National Academy of Sciences, Washington, D.C., 1989, pp. 115–124.

D. H. Hornig, U. Moser, and B. E. Glatthaar, in M. E. Shils and V. R. Young, eds., *Modern Nutrition in Health and Diseases*, 7th ed., Lea & Febiger, Philadelphia, 1988, pp. 417–435.

R. A. Jacob, J. H. Skala, and S. T. Omave, *Am. J. Clin. Nutrit.* **46**, 818–826 (1987).

R. Marcus and A. M. Coulston, in A. G. Gilman, L. S. Goodman, T. W. Rall, and F. Murari, eds., *Goodman and Gilman's Pharmacological Basis of Therapeutics*, 7th ed., Macmillan Publishing Co., New York, 1985, pp. 1151–1573.

U. Moser and A. Bendich, in L. J. Machlin, ed., *Handbook of Vitamins*, 2nd ed., Marcel Dekker, New York, 1991, pp. 195–232.

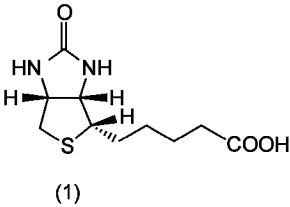
R. L. Pike and M. L. Brown, *Nutrition: An Integrated Approach*, 3rd ed., Macmillan Publishing Co., New York, 1986, pp. 131–136.
H. E. Sauberlich, in M. L. Brown, ed., *Present Knowledge in Nutrition*, 6th ed., International Life Sciences Institute, Nutrition Foundation, Washington, D.C., 1990, pp. 132–141.

Volker Kuellmer
Hoffmann-La Roche Inc.



BIOTIN

Biotin, [3a*S*-(3a α , 4 β , 6a α)]-hexahydro-2-oxo-1*H*-thieno[3,4-*d*]imidazole-4-pentanoic acid [58-85-5] (vitamin H, vitamin B₈, bios IIB, and coenzyme R) (1) is a water-soluble B complex vitamin. The name coenzyme R was coined during work on a protective factor, from the liver, for egg white injury. This protective factor was also called factor S, factor W or vitamin B_w (1–4). Biotin is a complex molecule having three stereocenters. There are eight stereoisomers of biotin; only the naturally occurring one is active in metabolism. The richest sources of biotin are yeast, liver, kidney, egg yolks, pancreas, and milk (5–7). The highest content of biotin in cow's milk occurs early in lactation. Plant materials, such as nuts, seeds, cereals such as oats and bulgar wheat, pollen, molasses, rice, soybeans, mushrooms, fresh vegetables such as cauliflower, split peas, cow peas, and legumes, and some fruits, are also good sources. In addition, small amounts of biotin are found in most fish, eg, mackerel, salmon, and sardines.



Isolation

In 1936, biotin was isolated from egg yolks (8), in 1939 from beef liver (9), and in 1942 from milk concentrates (10). Biotin-producing microorganisms exist in the large bowel but the extent and significance of this internal synthesis is unknown.

Biochemical Function

Biotin forms part of several enzyme systems and is necessary for normal growth and body function. Biotin functions as a cofactor for enzymes involved in carbon dioxide fixation and transfer. These reactions are important in the metabolism of carbohydrates, fats, and proteins, as well as promotion of the synthesis and formation of nicotinic acid, fatty acids, glycogen, and amino acids (5–7). Biotin is absorbed unchanged in the upper part of the small intestine and distributed to all tissues. Highest concentrations are found in the liver and kidneys. Little information is available on the transport and storage of biotin in humans or animals. A biotin level in urine of approximately 160 nmol 24 h or 70 nmol L, and a circulating level in blood, plasma, or serum of approximately 1500 pmol L seems to indicate an adequate supply of biotin for humans. However, reported levels for biotin in the blood and urine vary widely and are not a reliable indicator of nutritional status.

Nutritional Requirements

Since exact requirements for biotin are uncertain owing to incomplete knowledge regarding biotin availability from food and a lack of definitive studies concerning biotin requirements, the United States National Research Council has established a safe and adequate daily dietary level of intake for biotin rather than a recommended dietary allowance (RDA). The recommended daily intake of biotin in the United States for all persons ages seven years and older is 30–100 μ g d. In France and South Africa, a recommended daily intake for adults of up to 300 μ g d has been established, whereas in Singapore an intake of up to 400 μ g d is recommended. Diets consisting of a daily biotin intake of 28–42 μ g d are considered adequate. A level of 60 μ g d is sufficient for patients under long-term total parenteral nutrition (fed intravenously). An infant's daily intake ranges from 15–20 μ g d and is acquired mostly through human milk containing 3–20 μ g L, or formulas fortified with biotin. An adequate intake level of 10–30 μ d for infants and young children is recommended. The safe and adequate levels for daily dietary intake of biotin are listed in Table 1 (11). No side effects have been reported with oral doses of biotin as great as 40 mg or parenteral doses of 5–10 mg d in infants. No toxicity of biotin has been found (5–7).

Table 1. 1989 Estimated Safe and Adequate Daily Dietary Intake for Biotin

Group	Age, yr	Biotin, μ g
infants	0–0.5	10
	0.5–1.0	15
children	1–3	20
	4–6	25
	7–10	30
adolescents and adults	11 +	30–100

Physiological Significance

Animal. Dietary biotin deficiencies are extremely rare, perhaps because of the biosynthesis of biotin by intestinal microorganisms. However, biotin deficiency can be easily induced in most animals by ingestion of large amounts of raw egg white, which contain the biotin binding protein avidin. Avidin has a high affinity for biotin and binds with the ureido group to form a complex that is resistant to digestive enzymes. Biotin deficiency in animals causes a decrease in growth rate, loss of weight, alopecia, scaly dermatitis, hyperkeratosis, achromatrichia, and transverse fissures across the soft sole and cracks in the hard horn of the sole and claw wall. Minute amounts of biotin are known to be adequate to support body functions; therefore, biotin requirements for animals may be covered by the natural content of the feed and by the intestinal biosynthesis of biotin. However, biotin-responsive disease conditions not caused by primary biotin deficiency have been observed. One such condition is the fatty liver and kidney syndrome (FLKS). This syndrome has caused heavy economic losses in commercial broiler flocks. FLKS was found to be the result of a suboptimal biotin content in diet, coupled with other nutritional and environmental factors. Although the symptoms of FLKS are not those of classic biotin deficiency, they can be eliminated by biotin supplements. Supplementation has also been found to reduce the incidence and severity of claw lesions in pigs and weak hoof horn in horses (5–7) (see FEEDS AND FEED ADDITIVES).

Human. Biotin deficiencies in humans are extremely rare. The symptoms of deficiency are anorexia, fatigue, nausea, vomiting, hyperesthesia, glossitis, pallor, mental depression, dry scaly dermatitis, alopecia, and localized parasthesias. Both seborrheic dermatitis and Leiner's disease could be the signs of a biotin deficiency in infants (12). Several studies (5–7) have indicated that an erythematous exfoliative dermatitis is the first clinical sign of a biotin deficiency. Infants under six months of age and people who eat large quantities of raw egg whites are probably the groups most susceptible to biotin deficiency. Another susceptible group would be people who lack biotinidase. Biotinidase is the only enzyme capable of catalyzing the cleavage of biocytin (biotinyl- ϵ -lysine), the bound form of biotin. It has also been postulated that there may be a connection between biotin and the etiology of the sudden infant death syndrome (SIDS), which is a common cause of death in the first year of life. Unless biotin is included in their alimentation, patients receiving long-term parenteral nutrition are susceptible to biotin deficiency. All symptoms can be reversed and conditions corrected with biotin treatment. For adults, biotin dosage levels of 150–300 μg per day would be effective treatment. Finally, no drugs have been found to cause a potential biotin deficiency through short- or long-term use. However, low circulating biotin levels have been observed in heavy smokers, alcoholics, and patients under prolonged treatment with anticonvulsant drugs (5–7).

Chemical and Physical Properties

The empirical formula for biotin, $\text{C}_{10}\text{H}_{11}\text{N}_2\text{O}_3\text{S}$, was established in 1941 (13). The full structure of biotin was elucidated in 1942 by two independent groups (14–16). The first total chemical synthesis of biotin was achieved by Harris in 1945; this work confirmed the structure of biotin (17). The configuration of *d*-biotin was definitively established in 1956 by x-ray crystallographic analysis (18,19). The physical properties of *d*-biotin are found in Table 2. Chemically pure biotin is stable to air and heat. Biotin is also stable for months in mildly acidic and alkaline solutions; however, alkaline solutions, particularly above pH 9, are the least stable. Although biotin is not affected by reducing agents, it is incompatible with formaldehyde, chloramine T, oxidizing agents such as hydrogen peroxide and potassium permanganate, and strong acids such as nitrous acid (20). In most foodstuffs, biotin is bound to proteins, from which it is released in the intestine by protein hydrolysis and the enzyme biotinidase. Biotin is a highly stable, water-soluble vitamin that is resistant to most processing procedures, long-term storage, and normal cooking heat. Most of the biotin losses during cooking are the result of leaching into the cooking water. On the other hand, canning and food processing cause a moderate reduction in biotin content, owing to decomposition.

Table 2. Physical Properties of *d*-Biotin

Property	Characteristic
appearance	fine long needles white
color	
molecular weight	244.31
molecular formula	$\text{C}_{10}\text{H}_{11}\text{N}_2\text{O}_3\text{S}$
elemental analysis, wt %	
carbon	49.16
hydrogen	6.60
nitrogen	11.47
sulfur	13.12
melting point, °C	232–233
α^{D}_{21} , specific optical rotation, degrees	+89–91 ^a
dissociation constant, $\text{p}K_{\text{A}}$	6.3×10^{-6}
isoelectric point, pH	3.5
solubility, mg mL	
H ₂ O, RT	~0.22 ^b
95% alcohol, RT	~0.80
common organic solvents	insoluble

^a $c = 1$ in 0.1 N NaOH.

^b Higher in dil alkali.

Chemical Synthesis

Original Synthesis. The first attempted synthesis of *d*-biotin in 1945 afforded racemic biotin (Fig. 1). In this synthetic pathway, L-cysteine [52-90-4] (2) was converted to the methyl ester [5472-74-2] (3). An intramolecular Dieckmann condensation, during which stereochemical integrity was lost, was followed by decarboxylation to afford the thiophanone [57752-72-4] (4). Aldol condensation of the thiophanone with the aldehyde ester [6026-86-4] (5) afforded the conjugated thiophanone [85269-47-2] (6). The aldol condensation allowed for attachment of the properly functionalized C-5 side chain in one step, a weakness in many subsequent syntheses. The introduction of the second nitrogen was achieved by converting the keto group to the oxime [85269-48-3] (7), followed by zinc reduction to the thiophene [85269-49-4] (8). Stereoselective catalytic hydrogenation of the thiophene double bond gave mainly the thiophane [78763-60-7] (9), which was easily transformed to biotin. The nonselective catalytic hydrogenation and the epimerization that occurred during the conversion of (3) to (4) were the major factors leading to formation of racemic biotin (17).

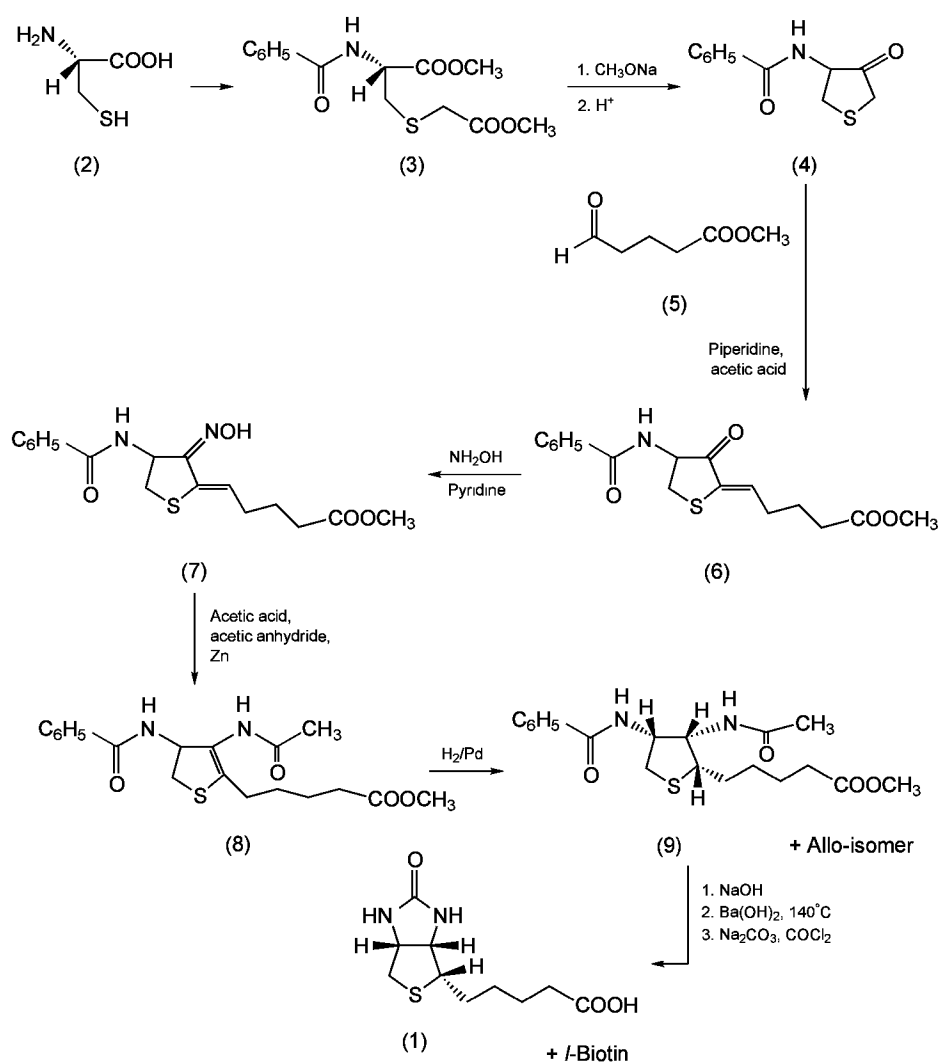
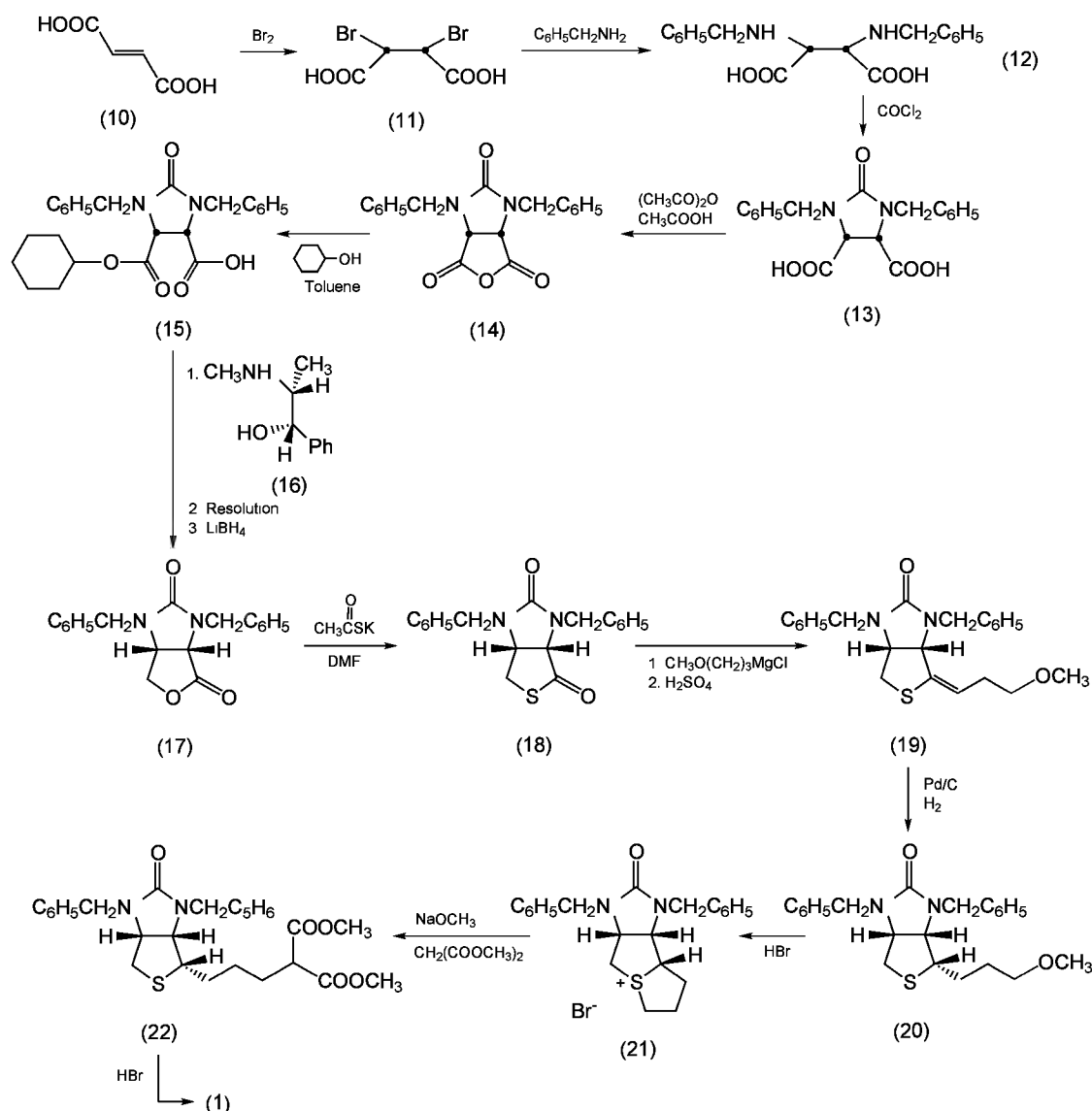


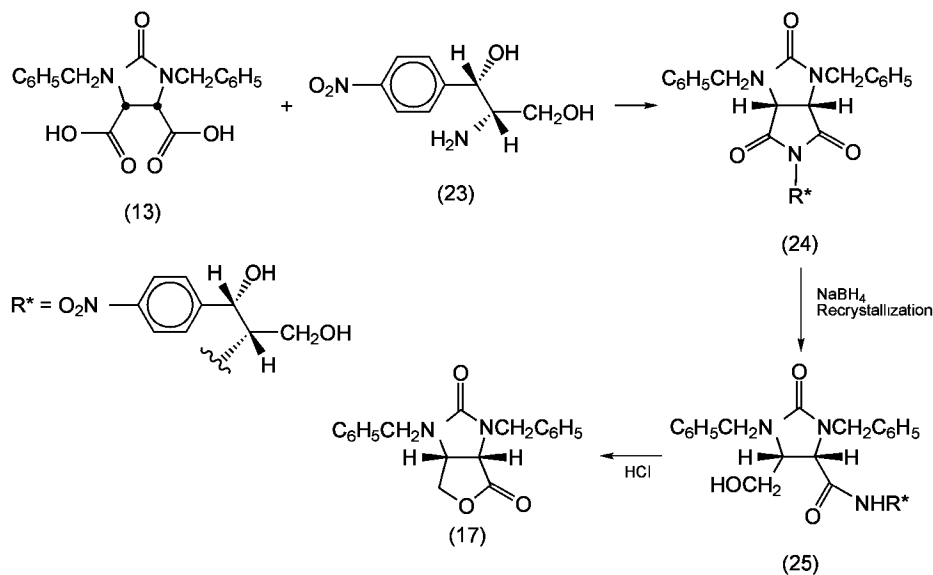
Fig. 1. Synthetic pathway for racemic biotin.

Original Asymmetric Synthesis. The efficient introduction of the three stereocenters in the all-*cis* configuration was first accomplished in the elegant synthesis developed by Sternbach (21–23) in 1949. This process, the Hoffmann-La Roche industrial synthesis of biotin, is still (ca 1997) the basis of industrial preparations. An improved version of the original Sternbach synthesis is shown in Figure 2 (24). The fixed *cis* position of the carboxyl groups of the imidazolidine dicarboxylic acid (cycloacid) (13) and the stereochemistry of the chiral carbons 4 and 5 of the imidazolidine ring, throughout the synthetic scheme, are established by the starting material, fumaric acid [110-17-8] (10). Fumaric acid is brominated to dibromosuccinic acid [608-36-6] (11) followed by reaction with benzylamine to give dibenzylaminosuccinic acid [55645-40-4] (12). Treatment with phosgene forms the imidazolone ring of cycloacid [51591-75-4] (13). Cycloacid is dehydrated with acetic anhydride to form cycloanhydride [56688-83-6] (14). Opening of cycloanhydride with cyclohexanol forms the racemic monoester (15). The mixture is resolved by salt formation with (+)-ephedrine [299-42-3] (16) in high yield. The desired *d*-ephedrine salt undergoes acid cleavage, followed by selective reduction of the ester functionality with lithium borohydride to form, after acidification, the *d*-lactone [56688-82-5] (17). The undesired *l*-ephedrine salt is recycled via the meso-cycloacid to cycloanhydride. The *d*-lactone, which possesses the desired configuration at two of the stereocenters, is converted to the *d*-thiolactone [56688-83-6] (18) using potassium thioacetate. The introduction of the C-5 side chain is accomplished by reacting the thiolactone with the Grignard reagent derived from 3-methoxypropyl chloride, followed by dehydration to give the thiophene [85611-62-7] (19). Stereoselective hydrogenation establishes the third stereocenter and affords the thiophene [33607-59-9] (20). The thiophene undergoes acid-catalyzed cyclization to form the key intermediate, *l*-thiophanium bromide [60209-10-1] (21). The last two carbons of the biotin side chain are added by reaction of thiophanium bromide with sodium dimethylmalonate to form the diester [8554-84-3] (22). Hydrolysis of the ester groups, decarboxylation and debenzylation using strong acid forms homochiral, pure *d*-biotin. Although a number of significant modifications to the Sternbach synthesis have appeared over the last 45 years, it is still the basis of today's industrial preparations.

Fig. 2. Synthetic pathway for *d*-biotin (Sternbach synthesis).

Synthetic Drawbacks. The major drawbacks in the Sternbach-Goldberg synthesis are the resolution-recycling of the intermediate that leads to *d*-lactone and the multiple manipulations required to add the five-carbon side chain. This sequence is inefficient, bringing with it a net loss of methanol, hydrogen bromide, carbon dioxide, and water that were once part of the molecule. In the resolution of the intermediate that leads to *d*-lactone, 50% of the product is the undesirable isomer, although this material is converted to the cycloacid for recycling. This is inherently inefficient and limits single-run production capacity by at least 50%.

Industrial Synthetic Improvements. One significant modification of the Sternbach process is the result of work by Sumitomo chemists in 1975, in which the optical resolution-reduction sequence is replaced with a more efficient asymmetric conversion of the *meso*-cycloacid (13) to the optically pure *d*-lactone (17) (Fig. 3) (25). The cycloacid is reacted with the optically active dihydroxyamine [2964-48-9] (23) to quantitatively yield the chiral imide [85317-83-5] (24). Diastereoselective reduction of the pro-*R*-carbonyl using sodium borohydride affords the optically pure hydroxyamide [85317-84-6] (25) after recrystallization. Acid hydrolysis of the amide then yields the desired *d*-lactone (17). A similar approach uses chiral alcohols to form diastereomeric half-esters stereoselectivity. These are reduced and directly converted to *d*-lactone (26). In both approaches, the desired diastereomeric half-amide or half-ester is formed in excess, thus avoiding the costly resolution step required in the Sternbach synthesis.

Fig. 3. Synthetic pathway for *d*-biotin (Sumitomo synthesis).

Another modification of the Sternbach method involves the direct attachment of the C-5 side chain to thiolactone or a functionalized thiolactone intermediate. One such method, possibly utilized by Sumitomo and reported several times by Lonza (27–29), involves the treatment of the thiolactone (14) with the five-carbon Wittig reagent (26) to give the olefin (27). Hydrogenation of the thiophene (27) followed by acid hydrolysis completes the synthesis of *d*-biotin. Another method that directly attaches the five-carbon side chain is patented by Lonza (Fig. 4). This method involves a Wittig reaction of a phosphonium salt derived from the thiolactone (14) and a C-5 aldehyde. The thiolactone (14) is reduced to the thiophanol (28), which is converted to the Wittig salt (29) with $\text{Ph}_3\text{P}^+\text{H}^- \text{BF}_3^-$. The ylide (29), in the presence of base, undergoes condensation with the aldehyde ester to form the thiophene (27), which in turn is catalytically hydrogenated to *d*-biotin.

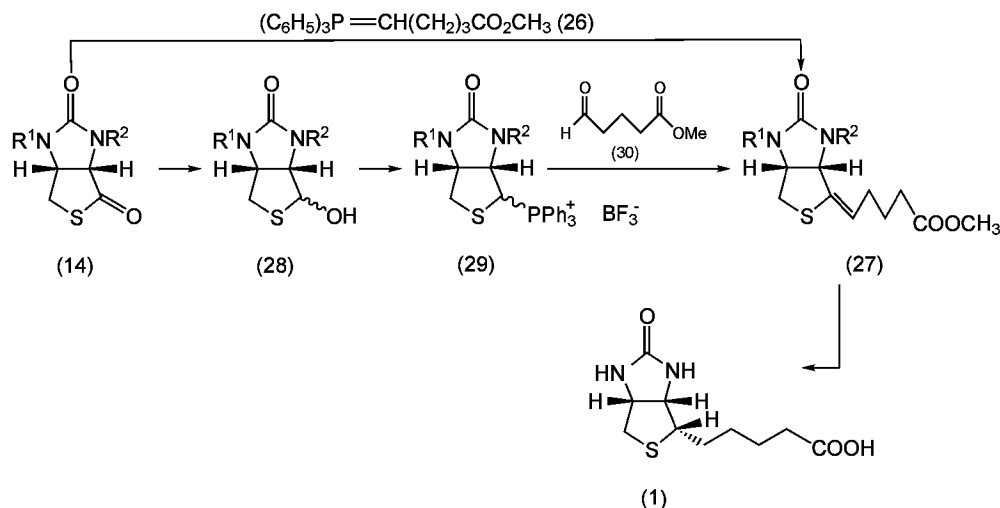


Fig. 4. Synthetic pathways for *d*-biotin (Lonza syntheses).

A final variation of the Sternbach method was reported by Lonza (27–29,31). This method (Fig. 5) not only involves direct attachment of the C-5 side chain to the lactone [118609-09-9] (33), but also introduces the key stereocenters early in the synthetic pathway via a chiral mono-protected imidazole intermediate [116291-87-3] (31). Catalytic diastereoselective hydrogenation of the furoimidazole (31) affords the *d*-lactone (33) as the major product, in 54% yield. The undesired diastereomeric *l*-lactone (32) can be easily separated by chromatography or crystallization. In the step following separation, the *d*-lactone is converted to the thiolactone [118609-15-7] (34) using a thiocarboxylic acid salt. The thiolactone undergoes a Grignard or Wittig reaction to incorporate the C-5 side chain. The thiophene [118609-16-8] (35) is then catalytically hydrogenated. This process hydrogenates the carbon–carbon double bond and also removes the chiral protecting group on the nitrogen. This process takes advantage of a short synthetic pathway and the use of a relatively inexpensive starting material. A recent modification of this Lonza process uses a chiral-substituted ferrocenyldisphosphine with a rhodium catalyst to stereoselectively reduce the enamine (31) to the *d*-lactone (33) in 99% yield. This improvement eliminates the need to separate the undesired *l*-lactone by chromatography or crystallization and increases the yield of the desired lactone from 54 to 99%, making the process extremely efficient (32,33).

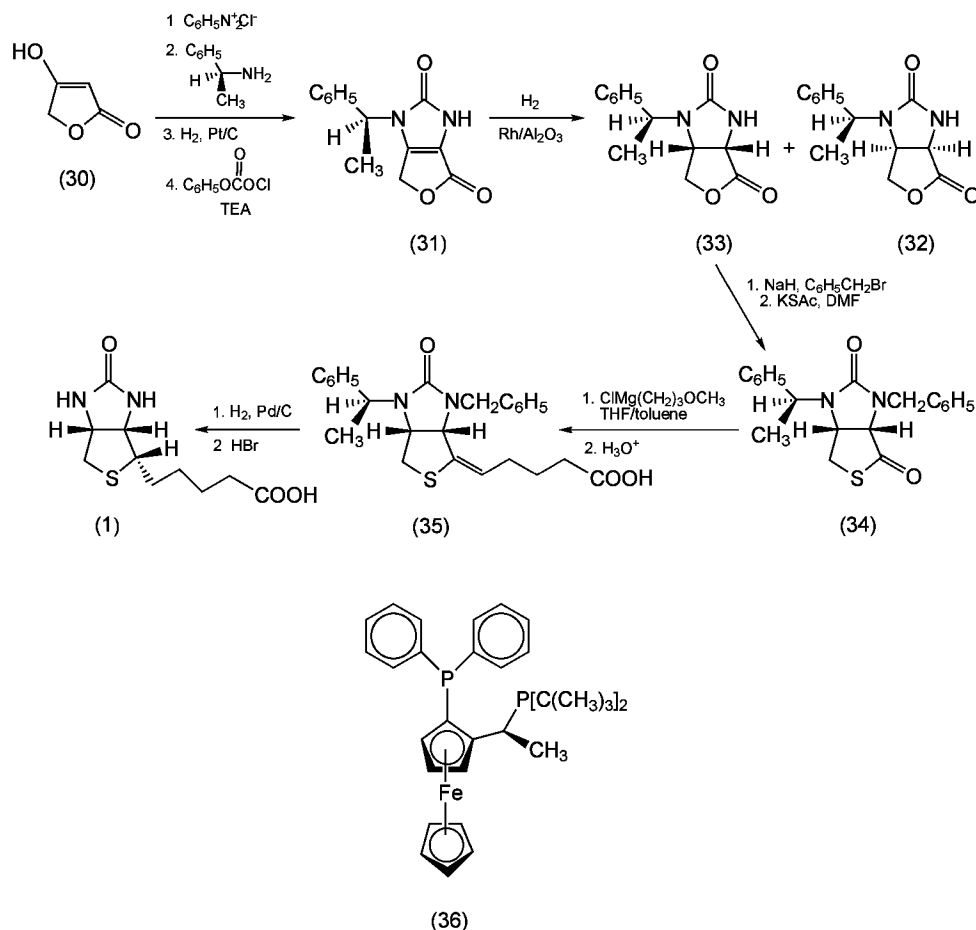


Fig. 5. Synthetic pathway for *d*-biotin (Lonza synthesis). An improved process uses the chiral ferrocenyldisphosphine (36) to introduce stereospecificity during the hydrogenation of lactone (31).

Novel Synthetic Methods. More recently, several novel syntheses of *d*-biotin, starting from a variety of chiral starting materials, have been

developed. Seven of these synthetic pathways start from L-cysteine (34–41), two from L-cystine (42,43), two from D-arabinose (44,45), and one from ribitol (46). Each of these methods has at least one of the following drawbacks: multiple steps with overall low yields; low yield steps associated with cyclization reactions; safety issues associated with reagents such as metal azides, methyl iodide, diazomethane, organoazide reagents, etc; environmental issues with reagents and by products such as triphenyl phosphine oxide and stannane salts; and costs of reagents such as 1,3-dicyclohexylcarbodiimide and trialkyltin hydride. Only one synthetic pathway starts from an achiral starting material, 2,5-dihydrothiophene-1,1-dioxide; this route requires a stereochemical resolution step (47–49). The drawbacks of this pathway are the commercially unavailable starting material and low conversions in several steps.

Biosynthesis

Biotin is produced by a multistep pathway in a variety of fungi, bacteria, and plants (50–56). The established pathway (50,56) in *E. coli* is shown in Figure 6. However, little is known about the initial steps that lead to pimelyl-CoA or of the mechanism of the transformation of desthiobiotin to biotin. Pimelic acid is believed to be the natural precursor of biotin for some microorganisms (51).

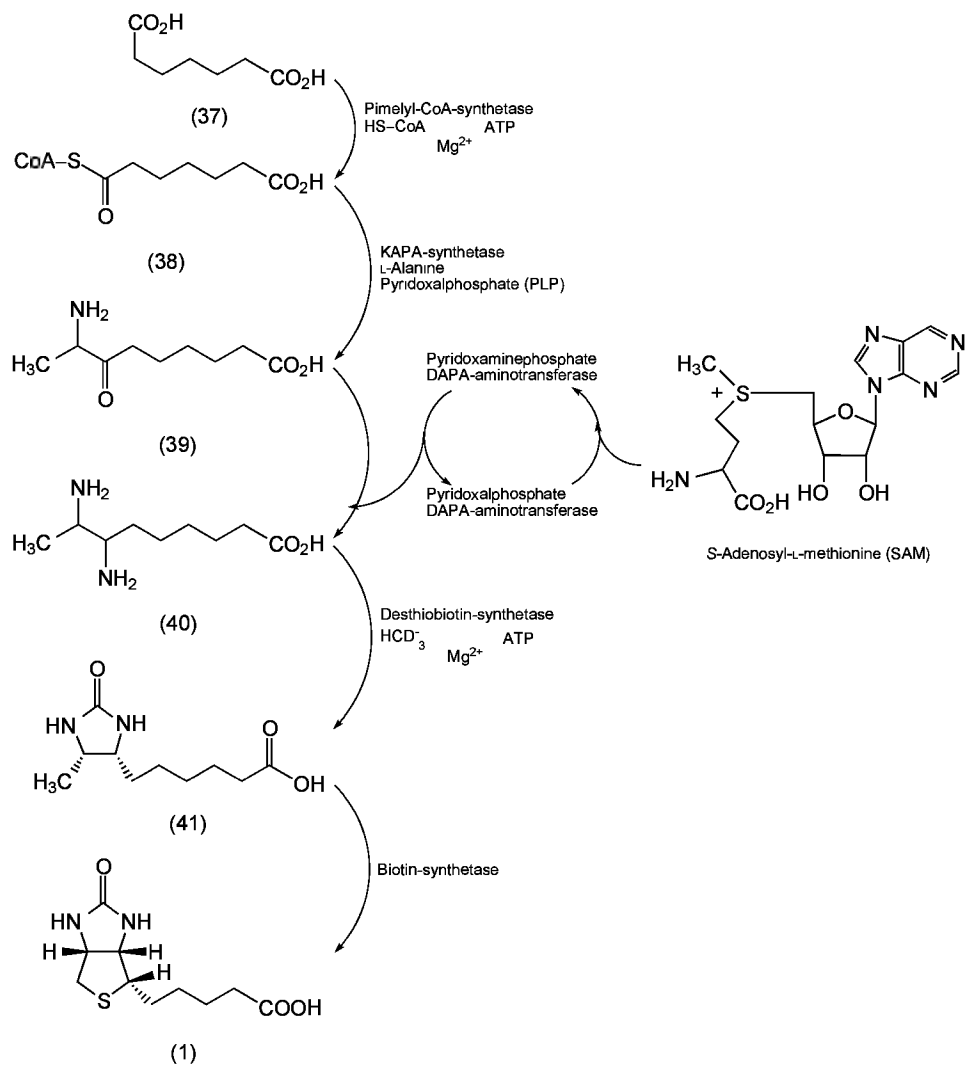


Fig. 6. Biosynthetic pathway of *d*-biotin.

Evidence that pimelic acid is a biotin precursor has been found in *Bacillus* species by feeding pimelic acid and observing the concomitant increase in biotin titers. Also, if labeled pimelic acid is used, labeled biotin is formed. In the biosynthetic pathway, pimelic acid [111-16-0] (37) is converted into pimelyl-CoA (38) by the enzyme pimelyl-CoA-synthetase in the presence of ATP and Mg²⁺ at 32°C and pH 7–8 (57). On the other hand, *E. coli* does not seem to rely on pimelic acid as a starting material for biotin synthesis. *E. coli* seems to form pimelyl-CoA by a pathway similar to that of fatty acid and polyketide synthesis (58). Pimelyl-CoA (38) is transformed into 7-keto-8-aminopelargonic acid [4707-58-8] (7-KAPA) (39) by reaction with L-alanine. The reaction requires pyridoxal-5'-phosphate (PLP) as a co-reactant and is catalyzed by the enzyme KAPA-synthetase (59). 7-Keto-8-amino-pelargonic acid (39) is converted into 7,8-diamino-pelargonic acid [157120-40-6] (DAPA) (40) by the enzyme DAPA-aminotransferase, which uses S-adenosyl-L-methionine (SAM) instead of a simple amino acid as the amino group donor (60,61). 7,8-Diaminopelargonic acid (40) is converted by the enzyme DTB-synthetase to desthiobiotin [533-48-2] (41). The optimal biological production of desthiobiotin requires ATP, Mg²⁺ and bicarbonate at 50°C and pH 7–8 (62–66). Finally desthiobiotin (41) is converted to biotin, supposedly by the enzyme biotin synthetase, which is encoded by the Bio B gene (56).

The conversion of desthiobiotin to biotin occurs in growing as well as resting cells in which the internal source of sulfur is believed to be cysteine or methionine. The introduction of sulfur into desthiobiotin has been investigated by several groups and it is proposed that the sulfur is introduced at C-4 of desthiobiotin with retention of configuration (67–69). The biosynthesis of biotin has led several groups to explore the feasibility of synthesizing biotin on a commercial scale by microorganisms. To date, no commercial scale total synthesis of biotin by fermentation is known. However, several microorganisms and mutant strains have been evaluated (Table 3). One of the problems associated with biotin biosynthesis via fermentation is that total biotin production decreases with increasing biotin concentrations in the fermentation broth. In fact, biotin strongly inhibits all steps of the biosynthesis except the synthesis of pimelyl-CoA. This inhibition by biotin leads to the isolation of several biotin vitamers. A vitamer is a compound, structurally similar to a vitamin, that exhibits varying degrees of vitamin activity. The most important biotin vitamer is desthiobiotin. Inhibition of biotin biosynthesis is the subject of much research.

Table 3. Organisms Used for Biotin Fermentation

Organism name	Titer, mg L	References
<i>Cornibacteria flavum</i>		70,71
<i>Brevibacterium flavum</i>	0.5 ^a	70,72

<i>Serratia marcescens</i>		73–78
	600	73
	500	74
	120	75,78
<i>Escherichia coli</i>	83	77
	2	79,80
<i>Bacillus sphaericus</i>		81–85
	70	81
	20	82
	365	83
<i>Rhodotorula rubra</i>		86
<i>Sporobolomyces roseus</i>		86
<i>Yarrowia lipolytica</i>		86
<i>Candida shehatae</i>		87
<i>Rhizopus delemar</i>	0.6	88,89

a

Analytical Methods

Biological Assay. Various analytical methods, including microbiological, biological, chemical, enzymatic and chromatographic assays, have been used to determine biotin levels in food, feed, and body fluids (79–81). The biological assay of biotin is conducted with rats or chicks made biotin-deficient by special diets containing either raw egg whites or avidin. The assay is based on growth response curves since test-animal weight gain is proportional to the logarithm of the biotin dose. A biological assay of biotin is advantageous because no pretreatment of the sample is necessary and the assay measures the amount of biotin available to the test animal. However, biological assays are time-consuming (four weeks to actually run the assay and six to seven weeks to prepare the biotin-deficient animals), costly to run, and require a large number of chicks or rats for accurate results.

Microbiological Assay. An alternative to the biological assay is the microbiological assay, which takes less time, costs less, requires less space, and has greater sensitivity than the biological assay. The typical microorganisms used for a microbiological assay of biotin are *Lactobacillus casei* ATCC 7469, *Neurospora crassa*, *Ochromonas danica*, *Saccharomyces cerevisiae* ATCC 7745, and *Lactobacillus plantarum* ATCC 8014. *Lactobacillus plantarum* ATCC 8014 is the test organism employed by most laboratories for biotin assay. Since biotin occurs in both the free and bound forms in nature and *L. plantarum* ATCC 8014 responds only to free biotin, all available biotin must be converted to free biotin prior to the microbiological assay. Bound biotin is usually converted to free biotin by acid hydrolysis with sulfuric acid or by digestion with papain. Hydrochloric acid should not be used in place of sulfuric acid for the acid hydrolysis because it may inactivate biotin. Lipids that stimulate the growth of *L. plantarum* interfere with the assay and have to be removed by filtration of the acid extracts or by preliminary ether extraction prior to the assay. A microbiological assay involves extraction, addition of graded levels of standard and sample to the assay tubes, addition of the medium, sterilization of the assay tubes, inoculation with *L. plantarum* and incubation. After the cell growth stops, the biotin content of the test material is determined by measuring the growth response of the organism (eg, colorimetrically, spectrophotometrically, or titrimetrically with sodium hydroxide) as compared to cell growth with known biotin concentration standards. Microbiological assay of biotin using *L. plantarum* is capable of detecting as little as 0.05 ng of biotin mL of test solution.

Isotope Dilution Assay. An isotope dilution assay for biotin, based on the high affinity of avidin for the ureido group of biotin, compares the binding of radioactive biotin and nonradioactive biotin with avidin. This method is sensitive to a level of 1–10 ng biotin (82–84), and the radiotracers typically used are [¹⁴C]biotin (83), [3H]biotin (84,85) or an ¹²⁵I-labeled biotin derivative (86). A variation of this approach uses ¹²⁵I-labeled avidin (87) for the assay.

High Performance Liquid Chromatography Analysis. The analysis of biotin has also been achieved by high performance liquid chromatography (hplc). Biotin has been analyzed in B-complex tablets, vitamin premixes, and multivitamin–multimineral preparations by reverse-phase, high performance liquid chromatographic methods (hplc) using a C¹⁸ column and uv detection at either 230 nm or 200 nm (88–91). Although this method can detect biotin in vitamin premixes, it is not sensitive enough to determine typical biotin levels in food or feed. Another method, one having greater sensitivity, is a reverse-phase ion-interaction reagent hplc method, with a biotin detection limit of 4 µg. An hplc method with even greater sensitivity involves derivatizing biotin prior to chromatographic separation. This technique can be used for either uv-absorbing derivatives, such as the bromoacetophenone ester (92), or fluorescent derivatives, such as methyl methoxycoumarin ester (92) and anthryldiazomethane ester (93). The biotin detection limit for the bromoacetophenone ester derivative at uv 254 nm was 50 ng. The detection limit for the methyl methoxycoumarin ester at excitation wavelength 360 nm and emission wavelength 410 nm was 5 ng; whereas, the limit of the anthryldiazomethane ester at excitation wavelength at 365 nm and emission wavelength 412 nm was 0.1–10 ng.

Gas Chromatography Analysis. From a sensitivity standpoint, a comparable technique is a gas chromatographic (gc) technique using flame ionization detection. This method has been used to quantify the trimethylsilyl ester derivative of biotin in agricultural premixes and pharmaceutical injectable preparations at detection limits of approximately 0.3 µg (94,95).

Specifications and Product Forms

According to the *Food Chemicals Codex* (96), the biologically active, food-grade form of biotin must have an assay of 97.5° o, a melting point in the range of 229–232°C, and a specific rotation at 25°C in 0.1 N NaOH in the range of + 89–93°. It must also contain less than 3 ppm of arsenic and less than 10 ppm of heavy metals, eg, lead, mercury, and copper. Finally, it must be able to be quantitatively sieved through U.S. Standard Sieves No. 80 using a mechanical shaker. Biotin is listed as GRAS in the *Code of Federal Regulations* (97,98) for use as a nutrient or dietary supplement (21 CFR 182.5159). The specifications for *d*-biotin for pharmaceutical applications are similar to those for food and are listed in the *United States Pharmacopeia* (99). No specifications for the optically inactive, *d,l*-biotin are given for either food or pharmaceutical applications.

Economic Aspects

The biotin market is divided between agricultural and human use, with ~90° o of biotin used in the animal health care market and ~10° o for the human nutritional market. The major producers of biotin are Hoffmann-La Roche, Lonza, E. Merck-Darmstadt, Rhône-Poulenc, Sumitomo Pharmaceutical, E. Sung, and Tanabe Seiyaku (100). Worldwide production of biotin in 1994 was approximately 60 metric tons. The list price for pure biotin in 1995 was ~\$7.00 g; whereas, the list price for technical feed-grade biotin was ~\$5.50 g. Biotin is used in various pharmaceutical, food, and special dietary products, including multivitamin preparations in liquid, tablet, capsule, or powder forms. One of the commercially available products of *d*-biotin is Brittit-1, which is a 1° o biotin titration used in food premixes.

BIBLIOGRAPHY

"Biotin" in *ECT* 1st ed., Vol. 2, pp. 519–525, by K. Hofmann, University of Pittsburgh; in *ECT* 2nd ed., Vol. 3, pp. 518–527, by J. D. Woodward, University of Reading; "Biotin" under "Vitamins" in *ECT* 3rd ed., Vol. 24, pp. 41–49, by M. R. Uskokovic, Hoffmann-La Roche Inc.

1. P. Gyngry and F.-W. Zilliken, in R. Amman and W. Dirsched, eds., *Fermente, Hormone, Vitamine*, Vol. 31, 3rd ed., Georg Thieme Verlag, Stuttgart, Germany, 1974, p. 766.

2. O. Isler, G. Bracher, S. Ghisla, and B. KLAUTLER, *Vitamine H. Wasserlyslische Vitamine*, Georg Thieme Verlag, Stuttgart, Germany, 1988, p. 231.

3. P. N. Achuta Murthy and S. P. Mistry, *Prog. Food Nutr. Sci.* **2**, 405 (1977).

4. F. A. Robinson, *The Vitamin Co-Factors of Enzyme Systems*, Pergamon Press, Oxford, U.K., 1966, p. 497.

5. *Encyclopedia of Food Science and Technology*, Pts. 1–8, Vol. 4, John Wiley & Sons, Inc., New York, 1991, p. 2764.

6. J. P. Bonjour, in L. Machlin, ed., *Handbook of Vitamins*, 2nd ed., Marcel Dekker Inc., New York, 1991, p. 393.

7. J. Gallagher, *Good Health with Vitamins and Minerals: A Complete Guide to a Lifetime of Safe and Effective Use*, Summit Books, New York, 1990, p. 67.

8. F. KUGL and B. TUNNIS, *Z. Physiol. Chem.* **242**, 43 (1936).

9. P. Gyurgy, R. Kuhn, and E. Lederer, *J. Biol. Chem.* **131**, 745 (1939).

10. D. B. Melville, K. Hofmann, E. Hague, and V. du Vigneaud, *J. Biol. Chem.* **142**, 615 (1942).

11. Food and Nutritional Board, National Research Council, *Recommended Dietary Allowances*, 10th ed., National Academy Press, Washington, D.C., 1989.

12. J. P. Bonjour, *Int. J. Vit. Nut. Res.* **47**, 107 (1977).

13. V. du Vigneaud, K. Hofmann, D. B. Melville, and J. R. Rachele, *J. Biol. Chem.* **140**, 763 (1941).

14. V. du Vigneaud and co-workers, *J. Biol. Chem.* **146**, 475 (1942).

15. D. B. Melville, A. W. Moyer, K. Hofmann, and V. du Vigneaud, *J. Biol. Chem.* **146**, 487 (1942).

16. P. Gyurgy and B. W. Langer, Jr., in W. H. Sebrell, Jr., R. S. Harris, eds., *The Vitamins*, 2nd ed., Academic Press, New York, 1968, p. 261.

17. S. A. Harris and co-workers, *JACS* **67**, 2096 (1945).

18. W. Traub, *Nature* **178**, 649 (1956).

19. J. Trotter and Y. A. Hamilton, *Biochemistry* **5**, 713 (1966).

20. S. Budavari, ed., *The Merck Index: An Encyclopedia of Chemicals, Drugs, and Biologicals*, 11th ed., Merck & Co., Inc., Rahway, N.J., 1989, p. 192.

21. **U.S. Pat. 2,489,232** (Nov. 22, 1949), M. W. Goldberg and L. H. Sternbach (to Hoffmann-La Roche Inc.).

22. **U.S. Pat. 2,489,235** (Nov. 22, 1949), M. W. Goldberg and L. H. Sternbach (to Hoffmann-La Roche Inc.).

23. **U.S. Pat. 2,489,238** (Nov. 22, 1949), M. W. Goldberg and L. H. Sternbach (to Hoffmann-La Roche Inc.).

24. M. Gerecke, J. P. Zimmermann, W. Aschwanden, *Helv. Chim. Acta*, **53**, 991 (1970).

25. **U.S. Pat. 3,876,656** (Apr. 8, 1975), A. Hisao, A. Yasuhiko, O. Shigeru, and S. Hiroyuki (to Sumitomo Chemical Co., Ltd.); Pat. Appl. E.P.O. 044,158 (Jan. 20, 1982), A. Hisao, A. Yasuhiko, O. Shigeru, and S. Hiroyuki (to Sumitomo Chemical Co., Ltd.).

26. **Eur. Pat. Appl.** EP 161,580 (Nov. 21, 1985), C. Wehli and H. Pauling (to F. Hoffmann-La Roche AG).

27. **Eur. Pat. Appl.** EP 273270-A1 (July 6, 1988), J. McGarrity and L. Tenud (to Lonza AG).

28. **Eur. Pat. Appl.** EP 270076-A1 (Feb. 12, 1986), M. Eyer, R. Fuchs, D. Laffan, J. F. McGarrity, T. Meul, and L. Tenud (to Lonza AG).

29. **U.S. Pat. 4,851,540** (Dec. 1, 1987), J. F. McGarrity, L. Tenud, and T. Meul (to Lonza AG).

30. **Eur. Pat. Appl.** EP 387,747 (Sept. 19, 1990), M. Eyer (to Lonza AG).

31. **U.S. Pat. 4,873,339** (Nov. 30, 1988), J. F. McGarrity, L. Tenud, and T. Meul (to Lonza AG).

32. **Eur. Pat. 624587** (Nov. 17, 1994), J. McGarrity, F. Spindler, R. Fuchs, and M. Eyer (to Lonza AG).

33. PCT Int. Appl. WO 94 24,137 (Oct. 27, 1994), M. Eyer and R. E. Merrill (to Lonza AG).

34. **U.S. Pat. 4,837,402** (Jan. 6, 1988), E. Poetsch and M. Casutt (to Merck Patent GmbH).

35. E. Poetsch and M. Casutt, *Chimia* **41**(5), 148 (1987).

36. Ger. Offen. DE 3,926,690 (Feb. 14, 1991), M. Casutt, E. Poetsch, and W. N. Spekamp (to Merck Patent GmbH).

37. F. D. Deroose and P. J. DeClercq, *Tetrahedron Lett.* **34**(27), 4365 (1993).

38. F. D. Deroose and P. J. DeClercq, *Tetrahedron Lett.* **35**(16), 2615 (1994).

39. F. D. Deroose and P. J. DeClercq, *J. Org. Chem.* **60**(2), 321 (1995).

40. T. Fujisawa, M. Nagai, Y. Koike, and M. Shimizu, *J. Org. Chem.*, **59**(20), 5865 (1994).

41. H. L. Lee, E. G. Baggiolini, and M. R. Uskokovic, *Tetrahedron*, **43**(21), 4887 (1987).

42. E. J. Corey and M. M. Mehrotra, *Tetrahedron Lett.* **29**(1), 57 (1988).

43. T. Ravindranathan, S. V. Hiremath, D. R. Reddy, and R. B. Tejwani, *Synth. Commun.* **18**(15), 1855 (1988).

44. Jpn. Kokai Tokkyo Koho, JP 61254590 (Nov. 12, 1986), T. Kono, Y. Shimakawa, M. Takahashi, T. Horisaki, and S. Masuda (to Teikoku Chemical Industry Co., Ltd.).

45. N. A. Hughes, K.-M. Kuhajda, and D. A. Miljkovic, *Carbohydr. Res.* **257**, 299 (1994).

46. D. A. Miljkovic and S. Velimirovic, *J. Serb. Chem. Soc.* **53**(1), 37 (1988).

47. H. A. Bates, L. Smilowitz, and J. Lin, *J. Org. Chem.* **50**, 899 (1985).

48. H. A. Bates, L. Smilowitz, and S. B. Rosenblum, *J. Chem. Soc. Chem. Commun.*, 353 (1985).

49. H. A. Bates and S. B. Rosenblum, *J. Org. Chem.* **51**, 3447 (1986).

50. M. Eisenberg, in F. Neidhardt, ed., *Escherichia coli and Samonella typhimzerium Cellular and Molecular Biology*, American Society for Microbiology, New York, 1987, p. 544.

51. R. Gloeckler and co-workers, *Gene* **87**, 63 (1990).

52. Y. Izumi and K. Ogata, *Adv. Appl. Microbiol.* **22**, 145 (1977).

53. P. Baldet, C. Alban, S. Axiotis, and R. Douce, *Arch. Biochem. Biophys.* **303**, 67 (1993).

54. T. Schneider, R. Dinkins, K. Robinson, J. Shellhammer, and D. Meinke, *Dev. Biol.* **131**, 161 (1989).

55. J. Shellhammer and D. Meinke, *Plant Physiol.* **93**, 1162 (1990).

56. M. A. Eisenberg, *Ann. N.Y. Acad. Sci.* **447**, 335 (1985).

57. O. Ploux, P. Soularue, A. Marquet, R. Gloeckler, and Y. Lemoine, *Biochem. J.* **287**, 685 (1992).

58. I. Sanyal, S.-L. Lee, and D. H. Flint, *JACS* **116**, 2637 (1994).

59. M. A. Eisenberg and C. Star, *J. Bacteriol.* **96**, 1291 (1968).

60. G. L. Stoner and M. A. Eisenberg, *J. Biol. Chem.* **250**, 4029 (1975).

61. C. H. Pai, *J. Bacteriol.* **105**, 793 (1971).

62. K. Krell and M. A. Eisenberg, *J. Biol. Chem.* **245**, 6558 (1970).

63. P. Cheeseman and C. H. Pai, *J. Bacteriol.* **104**, 726 (1970).

64. O. Ifuku, J. Kishimoto, S. Haze, M. Yanagi, and S. Fukushima, *Biosci. Biotechnol. Biochem.* **56**, 1780 (1992).

65. R. L. Baxter, A. J. Ramsey, L. A. McIver, and H. C. Baxter, *J. Chem. Soc., Chem. Commun.*, 559 (1994).

66. R. L. Baxter and H. C. Baxter, *J. Chem. Soc., Chem. Commun.*, 759 (1994).

67. F. Frappier, M. Jouany, A. Marquet, A. Olesker, and J.-C. Tabet, *J. Org. Chem.* **47**, 2257 (1982).

68. D. A. Trainer, R. J. Parry, and A. Gutterman, *JACS* **102**, 1467 (1980).

69. L. Even, D. Florentin, and A. Marquet, *Bull. Soc. Chim. Fr.* **127**, 758 (1990).

70. Jpn. Kokai Tokkyo Koho, JP06339371, (Dec. 13, 1994), Y. Yoneda, T. Abe, N. Hara, I. Ohsawa, and A. Fujisawa (to Nippon Zeon Co., Ltd.).

71. I. Ohsawa and co-workers, *J. Ferment. Bioen.*, **73**(2), 121 (1992).

72. **Eur. Pat. Appl.** EP 316229 (May 17, 1989), S. Haze, O. Ifuku, and J. Kishimoto (to Shiseido Co., Ltd.).

73. Jpn. Kokai Tokkyo Koho, JP 58060996 (Apr. 11, 1983), (to Nippon Zeon Co., Ltd.).

74. **Eur. Pat. Appl.** EP 375525 (June 27, 1990), R. Gloeckler, D. Speck, J. Sabatie, S. Brown, and Y. Lemoine (to Transgene S. A., Fr.).

75. B. M. Pearson, D. A. MacKenzie, and M. H. J. Keenan, *Lett. Appl. Microbiol.* **2**(2), 25 (1986).

76. J. C. Du Preez, J. L. F. Kock, A. M. T. Monteiro, and B. A. Prior, *FEMS Microbiol. Lett.* **28**(3), 271 (1985).

77. Otkrytiya, Izobret., *Prom. Obratzy, Tovarnye Znaki* (38), 93 (1983); Sov. Pat. SU 1047956 (Oct. 15, 1983), L. I. Vorob'eva, E. V. Shchelokova, and E. S. Naumova (to Moscow State University) (in Russian).

78. V. N. Maksimov, E. V. Shchelokova, and L. I. Vorob'eva, *Prikl. Biokhim. Mikrobiol.* **19**(3), 353 (1983).

79. I. D. Lumley and P. R. Lawrance, *J. Micronutr. Anal.*, **7**(3–4), 301 (1991).

80. J. Scheiner, J. Augustin, B. P. Klein, D. A. Becker, and P. B. Venugopal, eds., *Methods of Vitamin Assay*, 4th ed., John Wiley & Sons, Inc., New York, 1985, p. 535.

81. F. Frappier and M. Gaudry, in A. P. DeLeenheer, W. E. Lambert, and M. G. M. De Ruyter, eds., *Modern Chromatographic Analysis of the Vitamins*, Marcel Dekker, New York, 1985, p. 482.

82. Dakshinamurti, A. D. Landman, L. Ramamurti, and R. J. Constable, *Anal. Biochem.* **61**, 225 (1974).

83. R. L. Hood, *J. Sci. Food Agric.* **26**, 1847 (1975).

84. R. Rettenmaier, *Anal. Chim. Acta* **113**, 107 (1980).

85. T. Suomala and co-workers, *Clin. Chim. Acta* **177**, 253 (1988).

86. E. Livaniou, G. R. Evangelatos, and D. S. Ithakissions, *Clin. Chem.* **33**, 1983 (1987).

87. D. M. Mock and D. B. DuBois, *Anal. Biochem.* **155**, 272 (1986).

88. T. S. Hudson, S. Subramanian, and R. J. Allen, *J. Assoc. Official Anal. Chemists* **67**, 995 (1984).

89. S. L. Crivelli, P. F. Quirk, D. J. Steible, and S. P. Assenza, *Pharm. Res.* **4**, 261 (1987).

90. A. Rizzolo, C. Baldo, and A. Polesello, *J. Chromatogr.* **553**, 187 (1991).

91. M. C. Gennaro, *J. Chromatogr. Sci.* **29**, 410 (1991).

92. P. L. Desbene, S. Coustal, and F. Frappier, *Anal. Biochem.* **128**, 359 (1983).

93. K. Hayakawa and J. Oizumi, *J. Chromatogr.* **413**, 247 (1987).

94. V. Viswanathan, F. P. Mahn, V. S. Venturella, and B. Z. Senkowski, *J. Pharm. Sci.* **59**, 400 (1970).

95. H. Janecke and H. Voegelé, *Naturwissenschaften* **55**, 447 (1968).

96. Food and Nutrition Board, National Research Council, *Food Chemicals Codex*, 3rd ed., National Academy Press, Washington, D.C., 1981, p. 38.

97. *Evaluation of the Health Aspects of Biotin as a Food Ingredient*, Life Sciences Research Office, Federation of American Societies for Experimental Biology, Bethesda, Md., 1978.

98. *Code of Federal Regulations, Title 21, Food and Drugs, Parts 100–199 rev.*, Office of the Federal Register, General Services Administration, U.S. Government Printing Office, Washington, D.C., 1977.

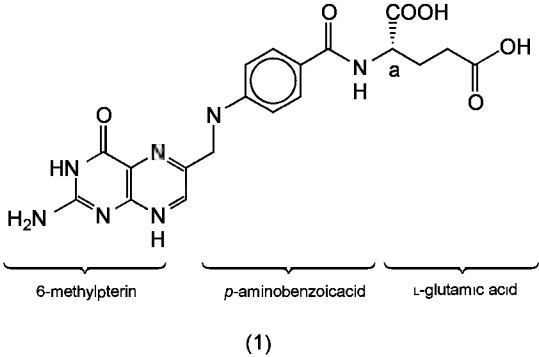
99. *The United States Pharmacopeia XXIII (USP XXIII–NF XVIII)*, United States Pharmacopeial Convention, Inc., Rockville, Md., 1995, p. 206.

100. J. D. Greer and B. Rhomberg, *CEH Marketing Research Report, Animals: Chemical Inputs for Nutrition and Health—Overview*, Chemical Economics Handbook-SRI International, Palo Alto, CA 1995, 201.8001A-B, Aug. 22, 1996.

Robert A. Outten
Hoffmann-La Roche Inc.

FOLIC ACID

Folic acid [59-30-3] (1) belongs to the group of B vitamins. The term folate is used to designate all members of the family of compounds based on the *N*-[(6-pteridinyl)methyl]-*p*-aminobenzoic acid skeleton conjugated with one or more L-glutamic acid units. In 1930, a dietary factor in yeast and crude liver extract was found to cure megaloblastic anemia in pregnant women (1). Purified folate was isolated by using two bioassay procedures, the microbiological growth assay and the assay for the antianemic factor for chickens (2–4). The growth factor from yeast and the antianemic factor (vitamin B₉) were later shown to be different entities. A crystalline compound was isolated from spinach, which was called folic acid (5,6). It was later shown that several of the above-mentioned factors belonged to the nutritionally and chemically related family of pteroylglutamic acid compounds. The structure of pteroylglutamic acid was elucidated in 1946 (7). The metabolically active forms of folic acid have a reduced pteridine ring and several glutamic acids residues. A detailed account on the discovery and early development of folic acid is available (8).



Occurrence, Source, and Bioavailability

Good food sources of folate are liver; fresh, dark green, leafy vegetables; beans; wheat germ; and yeasts (qv) (Table 1). Folic acid is synthesized only by microorganisms and plants (10–14). Most dietary folates exist in the polyglutamate form, which are converted to the more readily bioavailable monoglutamate form in the small intestine by the jejunal brush border folate conjugase. Certain foods such as cabbage and legumes contain conjugase inhibitors, which can decrease folate absorption.

Table 1. Select Contributors of Folate in the U.S. Diet^a

Ranking	Description	Total folate, %
1	orange juice	9.7
2	white bread, rolls, crackers	8.6
3	pinto, navy, other dried beans	7.1
4	green salad	6.8
5	cold cereals	5.0
6	eggs	4.6
9	liver	3.1
23	hamburger	1.2
25	spinach	1.0
30	green beans	0.8
34	broccoli	0.7

^a Ref. 9.

The total folate content of food varies, based on the method of preparation and length of storage (15–17). Different forms of folates occur in nature and the stability or bioavailability of each form varies. Most folates in food are easily oxidizable and therefore are susceptible to oxidation under aerobic conditions during storage and processing. Folic acid (commercial form) has superior bioavailability because it is more readily absorbed when compared to the tri- or heptaconjugates. The factors affecting the bioavailability of food folates are not well understood, but seem to include iron and vitamin C (qv) status. Deficiencies of both of these nutrients in humans is associated with impaired utilization of dietary folate. Improved research techniques such as use of radiolabeled folates has provided a powerful tool for determining bioavailability (18). Under fasting conditions, folic acid is almost completely absorbed, whereas only 50–80% of folyl polyglutamate is absorbed, as determined by urinary excretion measurements (19).

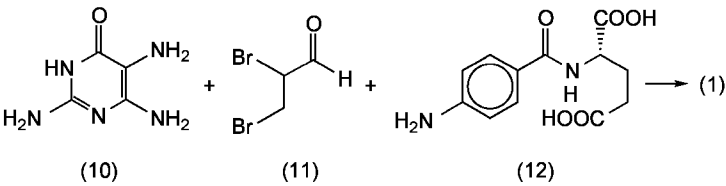
Chemical and Physical Properties

L-Folic acid (1) contains three subunits: 6-methylpterin, *p*-aminobenzoic acid, and L-glutamic acid. The *Chemical Abstracts* name is *N*-[4-[(2-amino-1,4-dihydro-4-oxo-6-pteridinyl)methyl]amino]benzoyl]-L-glutamic acid.

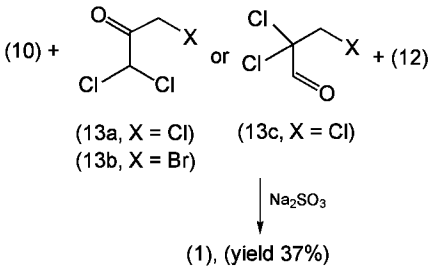
5,10-methenyltetrahydrofolic acid	(9)	[65981-89-7]	(pH 1) 352	23,900	$C_{20}H_{22}N_7O^+$
5,10-CH ⁺ H ₄ folate					(cation)

Synthesis

The first L-folic acid synthesis was based on the concept of a three-component, one-pot reaction (7,22). Triamino-4(3*H*)-pyrimidinone [1004-45-7] (10) was reacted simultaneously with C₃-dibromo aldehyde [5221-17-0] (11) and *p*-aminobenzoyl-L-glutamic acid [4271-30-1] (12) to yield folic acid (1).

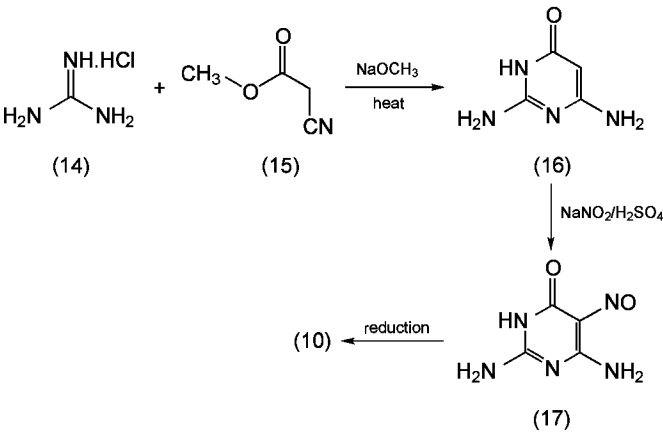


All known commercial syntheses are based on this approach with improvements in preparations of the three components (23). Shortly after the first synthesis, similar methods were published employing other C₃-halo compounds, such as 1,1,3-tribromo-2-propanone, 2,2,3-tribromopropanal (24), 2,2,3-trichloropropanal, and 1,1,3-trichloro-2-propanone (23).

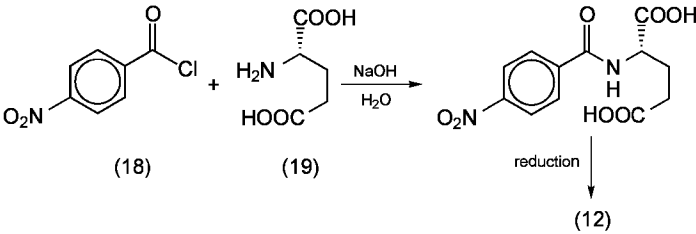


Yields were improved to >37% by the addition of sodium sulfite to the reaction mixture. Apart from the sulfite, the C₃-component unit has the greatest influence on the yield of folic acid. The use of nickel(II) chloride as an additive has been claimed to give higher yields (25).

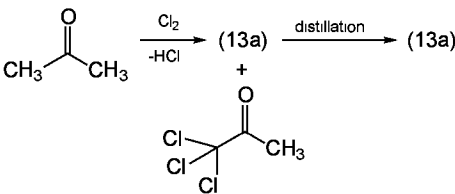
The required triamino-4(3*H*)-pyrimidinone is prepared in three steps starting from guanidine [50-01-1] (14) (26). Condensation with methylcyanoacetate [105-34-0] (15) under basic conditions, followed by nitrosation of the intermediate [56-06-4] (16), gives 2,6-diamino-5-nitrosopyrimidinone [2387-48-6] (17). Chemical reduction using sodium sulfite (27) or catalytic hydrogenation using Raney nickel (28) furnishes (10).



p-Aminobenzoyl-L-glutamic acid (12) is obtained by condensation of *p*-nitrobenzoyl chloride [122-04-3] (18) with L-glutamic acid [56-86-0] (19) under Schotten-Baumann conditions. This is followed by reduction of the nitro group with either sodium hydrogen sulfide (29) or by electrochemical methods (30).

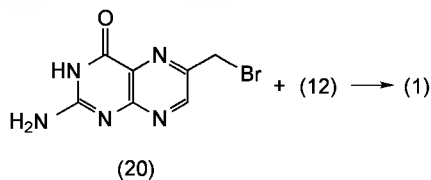


1,1,3-Trichloroacetone [921-03-9] (13a) is prepared by chlorination of acetone. The reaction is nonselective and the required compound is isolated by distillation. The selectivity has been improved by catalyzing the reaction with iodine (31).

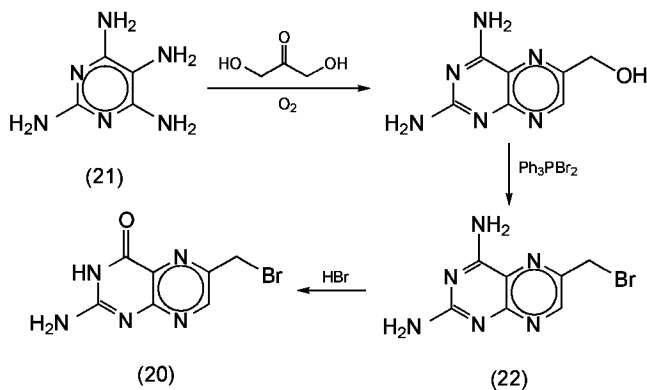


Alternative Approaches for Synthesis of L-Folic Acid. L-Folic acid (1) has been prepared in two steps by condensing

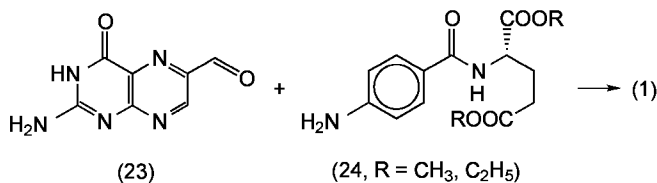
6-bromomethylpterin (20) with *p*-aminobenzoyl-L-glutamic acid (12) in 80° o yield (32). Dissolved folic acid further reacts easily with one more equivalent of 6-bromomethylpterin to form the undesired dialkylated aminobenzoyl-L-glutamic acid.



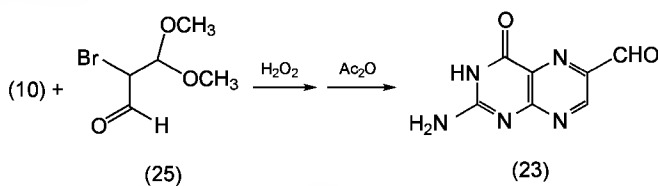
In spite of the good yields of L-folic acid obtained in this reaction, all of the published methods for the synthesis of 6-bromomethylpterin (20) are multistep procedures with low overall yields (33–36). For example, the route starting from 2,4,5,6-tetraaminopyrimidine [5392-28-9] (21) gave 6-bromomethylpterin (20) in three steps with an overall yield of only 18° o (33,35,36). This synthesis is not economical because the intermediate 6-bromomethyl-2,4-diamino-4-pterin (22) has to be deaminated in an additional step to form 6-bromomethylpterin (20).



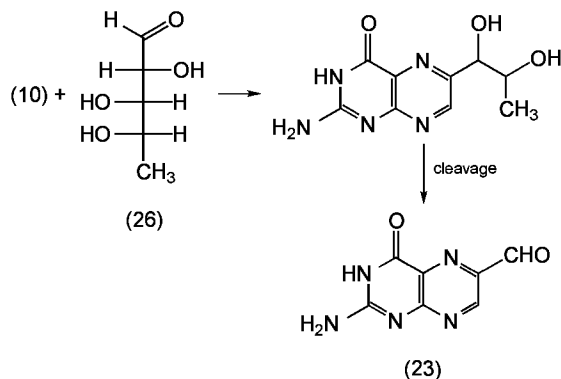
Another viable method for the synthesis of L-folic acid (1) starts from 6-formylpterin (23). The diester of L-glutamic acid (24) is condensed with 6-formylpterin (23). Reduction of the Schiff base with sodium borohydride is followed by hydrolysis to yield L-folic acid (37).



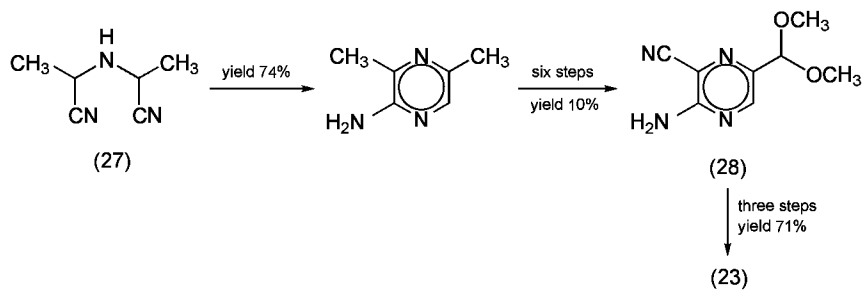
A cost-efficient synthesis of folic acid via Schiff base formation is feasible only if 6-formylpterin (23) is readily available. This compound is prepared by the reaction of 2-bromomalondialdehyde dimethylacetal [59453-00-8] (25) with triaminopyrimidinone (10), followed by acetylation and cleavage of the acetal to give compound (23) in 51° o overall yield (38).



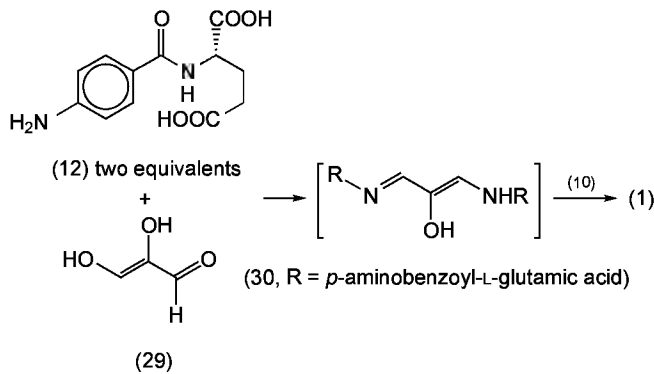
A second approach for the synthesis of 6-formylpterin (23) involves the condensation of triaminopyrimidinone (10) with 5-deoxy-L-arabinose (26). The key diol is obtained in four steps starting from compound (10). Cleavage of the diol side chain is achieved either with periodate (39) or with lead(IV) (40) to furnish 6-formylpterin (23) in 45° o overall yield.



A third approach for the synthesis of 6-formylpterin (23) starts from iminodipropionitrile [2869-25-2] (27). The intermediate pyrazine (28) is also prepared starting from chloropyruvaldehyde oxime (41,42). The required formylpterin (23) is obtained in three steps in 71° o yield, starting from the intermediate pyrazine (28). A few other routes for the synthesis of 6-formylpterin (23) are described in the literature. All are multistep procedures with only moderate overall yields (43–45).



A new variant of the three-component, one-pot synthesis of L-folic acid has been reported by Hoffmann-La Roche Inc.



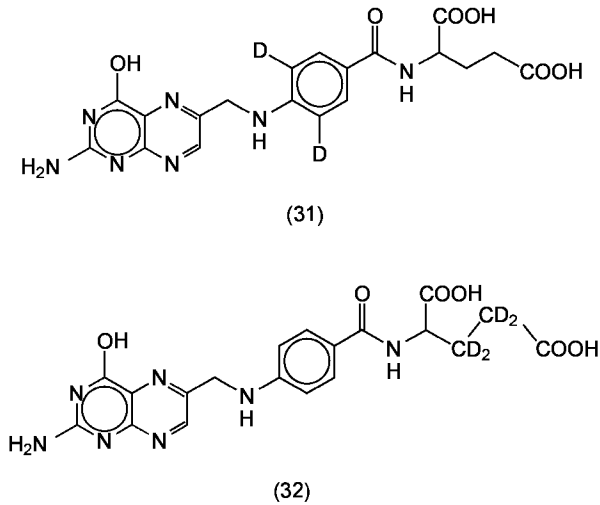
One equivalent of 2-hydroxymalondialdehyde [497-15-4] (29) is condensed with two equivalents of *p*-aminobenzoyl-L-glutamic acid (12). The intermediate dimine (30) is treated with one equivalent of triaminopyrimidinone (10) to obtain L-folic acid in 84% yield (46).

Fermentation

Development of an economically viable production process for folic acid either by genetically engineered microorganisms or by extraction from natural sources is not yet feasible.

Labeled Compounds

Radiolabeled folate provides a powerful tool for folate bioavailability studies in animals and for diagnostic procedures in humans. Deuteration at the 3- and 5-positions of the central benzene ring of folic acid (31) was accomplished by catalytic debromination (47,48) or acid-catalyzed exchange reaction (49). Alternatively, deuterium-labeled folic acid (32) was prepared by condensing pteric acid with commercially available labeled glutamic acid (50).



Derivatives and Analogues

The metabolically active H₄ folate cofactors (see Table 2) are prepared synthetically as follows. 7,8-Dihydrofolic acid (2) and 5,6,7,8-tetrahydrofolic acid (3) are prepared via catalytic hydrogenation of folic acid under controlled reaction conditions (51,52). Optical rotation of (6*S*,*R*)-H₄ folate (3) is [α]_D²⁷ +14.9° (0.1 N NaOH) and the natural (6*S*)-H₄ folate (3) is [α]_D²⁷ 16.9° (0.1 N NaOH).

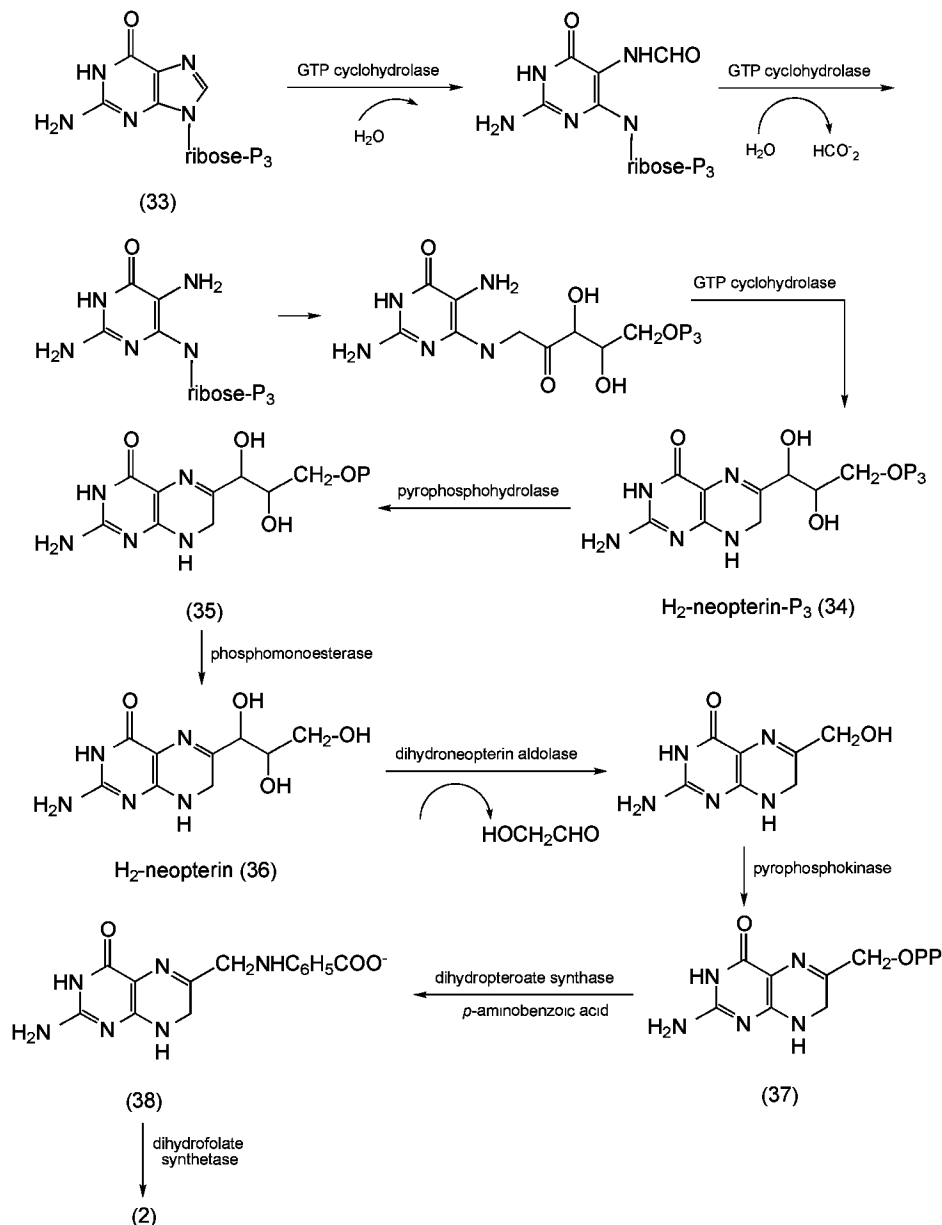
5-Methyltetrahydrofolic acid (5-CH₃-H₄ folate) (4) is involved in methionine biosynthesis. Condensation of formaldehyde with H₄ folate (3), followed by the reduction of the intermediate 5,10-CH₂-H₄ folate (5) with sodium borohydride gave 5-CH₃-H₄ folate (4) (53). 5,10-Methylenetetrahydrofolic acid (5,10-CH₂-H₄ folate) (5) is a coenzyme in thymidylate biosynthesis; the natural (6*R*)-stereoisomer is prepared by enzymatic reduction of H₂ folate (2), followed by condensation with formaldehyde (54).

Formylation of H₄ folate (3) or hydrolysis of 5, 10 - CH₂ - H₄ folate (9) gives (6*R*,*S*)-5-formyltetrahydrofolic acid (6) (5-HCO-H₄ folate) (55). On the other hand, (6*S*)-5-HCO-H₄ folate is obtained by selective crystallization in the form of its calcium salt from the diastereomeric mixture of (6*S*,*R*)-5-HCO-H₄ folate (56). 10-Formyltetrahydrofolic acid (7) is a coenzyme in purine synthesis which is synthesized by hydrolysis of 5, 10 - CH₂ - H₄ folate (9) or by hydrogenation of 10-CHO-folate (57).

Folic acid analogues containing amino acids other than glutamate, and also folate covalently bound to a protein for the purposes of antibody production, have been prepared (58–60). Methotrexate is an analogue of folic acid that is widely used in cancer chemotherapy (61) (see CHEMOTHERAPEUTICS, ANTICANCER). Other analogues such as trimethoprim and pyrimethamine are used in the treatment of malaria and protozoal diseases (62). These analogues bind extremely tightly to dihydrofolate reductase.

Biosynthesis

Folic acid is synthesized both in microorganisms and in plants. Guanosine-5-triphosphate (GTP) (33), *p*-aminobenzoic acid (PABA), and L-glutamic acid are the precursors. Reviews are available for details (63,64). The sequence of reactions responsible for the enzymatic conversion of GTP to 7,8-dihydrofolic acid (2) is shown.



In *E. coli*, GTP cyclohydrolase catalyzes the conversion of GTP (33) into 7,8-dihydroneopterin triphosphate (34) via a three-step sequence. Hydrolysis of the triphosphate group of (34) is achieved by a nonspecific pyrophosphatase to afford dihydroneopterin (35) (65). The free alcohol (36) is obtained by the removal of residual phosphate by an unknown phosphomonoesterase. The dihydroneopterin undergoes a retro-aldol reaction with the elimination of a hydroxy acetaldehyde moiety. Addition of a pyrophosphate group affords hydroxymethyl-7,8-dihydropterin pyrophosphate (37). Dihydropteroate synthase catalyzes the condensation of hydroxymethyl-7,8-dihydropteroyl pyrophosphate with PABA to furnish 7,8-dihydropteroylglutamate (38). Finally, L-glutamic acid is condensed with 7,8-dihydropteroylglutamate in the presence of dihydrofolate synthetase.

Analytical Methods

Analysis of folic acid is difficult because most natural folates exist in the polyglutamate form and there is variation in the oxidation state of the single-carbon substituent. Determination of the individual folate vitamers is complicated; assay simplification is achieved by determining folate in the monoglutamyl or diglutamyl forms after enzymatic deconjugation. Methods for determining folic acid in food and feed include biological, microbiological, chemical, chromatographic, and radiometric assays. The microbiological assay using *Lactobacillus casei* is the official method of the Association of Official Analytical Chemists (AOAC). Folyl polyglutamates react very slowly with this organism compared to mono- and diglutamates. As a result, it is required to hydrolyze the polyglutamate chain using γ -glutamylhydrolase prior to microbiological analysis (66). The monoglutamate in the food and feed extract is separated using anion-exchange column chromatography, followed by differential assay of the separated fraction by a microbiological assay.

A radioassay procedure has been developed to determine folic acid in erythrocyte and blood samples. The method is based on competitive protein binding between radiolabeled and unlabeled folate compounds for folic acid binding protein. A very sensitive, nonisotopic microtitration plate, folate-binding protein assay (FBPA) was developed to measure down to the 6 pg level of folyl monoglutamate and vitamin-active folate (67). A hplc method is useful for analysis of high potency premix samples containing folic acid, along with other water-soluble vitamins. Ion-pair reverse-phase columns using uv detection at 280 nm or post-column derivatization techniques have been employed to determine free folic acid. Details on hplc applications are available (68–70).

Deficiency

Folic acid is a precursor of several important enzyme cofactors required for the synthesis of nucleic acids (qv) and the metabolism of certain amino acids. Folic acid deficiency results in an inability to produce deoxyribonucleic acid (DNA), ribonucleic acid (RNA), and certain proteins (qv). Megaloblastic anemia is a common symptom of folate deficiency owing to rapid red blood cell turnover and the high metabolic requirement of hematopoietic tissue. One of the clinical signs of acute folate deficiency includes a red and painful tongue. Vitamin B₁₂ and folate share a common metabolic pathway, the methionine synthase reaction. Therefore a differential diagnosis is required to measure folic acid deficiency because both folic acid and vitamin B₁₂ deficiency cause

megaloblastic anemia. Serum and red blood cell levels of folate are measured to confirm and diagnose the deficiency. Serum folate levels less than 3 ng mL are diagnostic of a deficiency. Some forms of dietary folic acid are more poorly absorbed by the elderly than by younger individuals. However, vitamin supplements containing the monoglutamate form are well absorbed in all age groups (71).

Folate antagonists (eg, methotrexate and certain antiepileptics) are used in treatment for various diseases, but their administration can lead to a functional folate deficiency. Folate utilization can be impaired by a depletion of zinc (see ZINC COMPOUNDS). In humans, the intestinal brush border folate conjugase is a zinc metalloenzyme (72). One study indicates that the substantial consumption of alcohol, when combined with an inadequate intake of folate and methionine, may increase the risk of colon cancer (73). Based on this study, it is recommended to avoid excess alcohol consumption and increase folate intake to lower the risk of colon cancer.

Incomplete closure of neural tube during the embryo development in humans can lead to spina bifida. The condition is characterized by an opening in the spinal cord and results in physical disability in a child. Incomplete closure of the skull produces anencephaly. These and similar conditions are collectively called neutral tube defects (NTD). Each year in the United States approximately 2500 infants are born with spina bifida and anencephaly and an estimated 1500 fetuses affected by these birth defects are spontaneously aborted (74). It has been shown that folic acid given at 400 µg day prevents the recurrence of NTD and that doses of 800 µg day prevent both the occurrence and recurrence of NTD in the majority of cases. Published studies also indicate that folic acid supplementation taken 6 weeks before conception may reduce the risk of neural tube defects by at least 50% (74,75). Folic acid may be more effective in reducing neural tube defect incidence than conjugated food folate because free folic acid is more readily absorbed (76). A cost-benefit analysis, based on the U.S. population, of preventable neural tube defects indicates that folic acid fortification of grains in the United States may yield a substantial economic benefit (77). The U.S. Food and Drug Administration (FDA) has amended the standards of identity for several enriched grain products. The agency is requiring that these products be fortified with folic acid at levels ranging from 0.95 to 3.09 mg per kg of product (78). This is the first B-vitamin fortification requirement since 1943 when the U.S. government mandated fortification of flour with niacin, thiamin, and riboflavin. Incidence of cleft lip cleft palate has been reported in animal studies due to folic acid deficiency. Poor folic acid status has been associated with megaloblastic changes in the cells of the uterine, cervix (79), and intestinal epithelium (80).

Homocysteine arises from dietary methionine. High levels of homocysteine (hyperhomocysteinemia) are a risk factor for occlusive vascular diseases including atherosclerosis and thrombosis (81–84). In a controlled study, serum folate concentrations of <9.2 nmol L were linked with elevated levels of plasma homocysteine. Elevated homocysteine levels have been associated also with ischemic stroke (9). The mechanism by which high levels of homocysteine produce vascular damage are, as of yet, not completely understood. Interaction of homocysteine with platelets or endothelial cells has been proposed as a possible mechanism. Clinically, homocysteine levels can be lowered by administration of vitamin B , vitamin B₁₂, and folic acid.

Requirements

The amount of folic acid required for daily intake is estimated based on the minimum amount required to maintain a certain level of serum folate. The recommended dietary allowance (RDA) for folic acid accounts for daily losses and makes allowances for variation in individual needs and bioavailability from food sources (85). The U.S. recommended daily allowance for adults is 400 µg and for pregnant women is 800 µ (Table 4).

Table 4. RDA and U.S. RDA for Folic Acid^a

Group	Age	Folic acid, µg
RDA		
infants	0–0.5	25
	0.5–1.0	35
children	1–3	50
	4–6	75
	7–10	100
males	11–14	150
	>15–51	200
females	11–14	150
	15–51	180
pregnant, lactating		400
1st six months		280
2nd six months		260
U.S. RDA		
infants <13 months		100
children <4 yr		200
adults and children >4 yr		400
pregnant or lactating women		800

^a Ref. 21.

Animal Nutrition

To obtain optimal performance of farm animals, folic acid supplementation is required (86) and as is the case with most of the vitamins, the majority of worldwide consumption is as feed supplements. The folic acid requirement for chickens and pigs is about 0.2–0.5 mg of folic acid kg diet and 0.3 mg kg diet, respectively. Increased amounts, 0.5–1.0 mg kg feed for chickens and 0.5–2.0 mg kg for swine, are recommended under commercial production conditions (87). The degree of intestinal folic acid synthesis and the utilization by the animal dictates the folic acid requirements for monogastric species. Also, the self-synthesis of folacin is dependent on dietary composition (88).

Folacin requirements are related to the type and level of production. The more rapid the growth or production rates, the greater the need for folacin owing to its role in DNA synthesis. In poultry, the requirement for egg hatchability is higher than for production (88). In swine, folic acid supplementation has been shown to increase fertility and growth rates (89).

Metabolism

The principal function of the folate coenzyme is to carry one-carbon units. Dietary folylpolyglutamate is hydrolyzed to the monoglutamate form by folyl polyglutamate hydrolyases (conjugases) prior to transport across the intestinal mucosa. Intestinal folate absorption occurs in the jejunum. A slightly acidic pH of 6 is optimum for intestinal absorption and transportation into the blood stream (90). It was shown by the competitive inhibition method (91) that a single protein seems to be responsible for transportation of the monoglutamate. Transportation may occur by an anion-exchange mechanism (92). Folate transportation may be partially dependent on sodium ions in the cell medium (93).

Three forms of folate appear to be transported in the blood: folic acid, folate loosely bound to low affinity binder serum proteins (such as albumin, α-macroglobulin, and transferrin), and folate bound to high affinity protein binders. Approximately 5% of total serum folate is being transported by high

affinity protein binders but the function of these proteins is not well understood (94,95). Folic acid is stored in folyl polyglutamate form. The liver and other tissues (mitochondria) convert folic acid and methyltetrahydrofolate into folylpolyglutamate by employing polyglutamate synthetase. The total folate pool in adult humans is estimated to be around 5–10 mg, with half of this in the liver; folylpolyglutamate synthetase activity is highest in liver.

The metabolism of folic acid involves reduction of the pterin ring to different forms of tetrahydrofolylglutamate. The reduction is catalyzed by dihydrofolate reductase and NADPH functions as a hydrogen donor. The metabolic roles of the folate coenzymes are to serve as acceptors or donors of one-carbon units in a variety of reactions. These one-carbon units exist in different oxidation states and include methanol, formaldehyde, and formate. The resulting tetrahydrofolylglutamate is an enzyme cofactor in amino acid metabolism and in the biosynthesis of purine and pyrimidines (10,96). The one-carbon unit is attached at either the N-5 or N-10 position. The activated one-carbon unit of 5,10-methylene-H₄ folate (5) is a substrate of T-synthase, an important enzyme of growing cells. 5-10-Methylene-H₄ folate (5) is reduced to 5-methyl-H₄ folate (4) and is used in methionine biosynthesis. Alternatively, it can be oxidized to 10-formyl-H₄ folate (7) for use in the purine biosynthetic pathway.

Toxicity

Folic acid is safe, even at levels of daily oral supplementation up to 5–10 mg (97). Gastrointestinal upset and an altered sleep pattern have been reported at 15 mg day (98). A high intake of folic acid can mask the clinical signs of pernicious anemia which results from vitamin B₁₂ deficiency and recurrence of epilepsy in epileptics treated with drugs with antifolate activity (99). The acute toxicity (LD₅₀) is approximately 500 and 600 mg per kg body weight for rats and mice, respectively (100).

Uses

L-Folic acid is available as a crystalline dihydrate containing 8° water. Approximately 80° of the commercial production is consumed for feed enrichment in animal nutrition. Folic acid is being offered by the pharmaceutical industry for therapeutic and prophylactic use (see PHARMACEUTICALS). Pharmacological doses of folic acid are commonly used as a rescue dose during cancer chemotherapy, in women using oral contraceptives, and alcoholics. Several studies have provided evidence that multivitamins or folic acid (0.8–4 mg day) supplementation prevent the majority of neural tube defects (101).

Economic Aspects

The world production of synthetic L-folic acid 1996 was estimated at 400 metric tons per year. The total market is expected to grow with increasing recognition of need, especially during pregnancy and lactation. The principal producers of folic acid are Hoffmann-La Roche, Takeda, Sumika Fine Chemical (previously Yodogawa Pharmaceuticals), Kongo, and three Chinese companies, Changzhou Pharmaceuticals, Changshu Hugang Pharmaceuticals, and Zhejiang Jiangnan Pharmaceuticals. Smaller quantities are also produced by companies in India, China, and Russia. The 1996 sale price varied between \$50 to \$130 kg.

Conclusions

All known commercial syntheses of folic acid are based on the three-component process developed in the late 1940s. Industrial production of folic acid by genetically engineered microorganisms or extraction from natural sources is not yet economically viable. The mechanism governing folate turnover and excretion is still poorly understood. Improved research techniques will aid in the development of a better understanding of factors affecting intestinal absorption and *in vivo* kinetics in human beings. Folic acid and multivitamin supplement use are associated with a decreased occurrence of neural tube defects. Abnormalities in homocysteine metabolism are observed in many women who have given birth to children with neural tube defects and in individuals with cardiovascular disease. Folic acid is likely to be involved in overcoming this abnormality. The exact mechanism by which folic acid prevents these diseases is a current active area of research.

BIBLIOGRAPHY

"Folic Acid," in *ECT* 2nd ed., Vol. 6, pp. 778–785, by M. E. Hultquist and T. H. Jukes.

1. L. Wills, P. W. Clutterbuck, and P. D. F. Evans, *Biochem. J.* **31**, 2136 (1937).
2. A. G. Hogan and E. M. Parrott, *J. Biol. Chem.* **132**, 507 (1940).
3. E. E. Snell and W. H. Peterson, *J. Bacteriol.* **39**, 273 (1940).
4. H. K. Mitchell, E. E. Snell, and R. J. Williams, *J. Am. Chem. Soc.* **63**, 2284 (1941).
5. S. B. Binkley, O. D. Bird, E. S. Bloom, R. A. Brown, D. G. Calkins, C. J. Campbell, A. D. Emmett, and J. J. Pfiffner, *Science*, **100**, 36 (1944).
6. J. J. Pfiffner, D. G. Calkins, E. S. Bloom, and L. B. O'Dell, *J. Am. Chem. Soc.* **68**, 1392 (1946).
7. R. B. Angier and co-workers, *Science*, **103**, 667 (1946).
8. A. D. Welch, *Perspect. Biol. Med.* **27**, 64 (1983).
9. W. H. Giles, S. J. Kittner, R. F. Anda, J. B. Croft, and M. L. Casper, *Stroke*, **26**, 1166 (1995).
10. W. Friedrich, *Handbuch der vitamine*, Urban & Schwarzenberg, Baltimore, Md., 1987, p. 398.
11. B. Botticher and R. Kluthe, in K. Pietrzik, ed., *Folsäure-Mangel*, W. Zuckschwerdt Verlag, Germany, 1987, p. 15.
12. T. Brody, in L. J. Machlin, ed., *Handbook of Vitamins*, Marcel Dekker Inc., New York, 1991, p. 453.
13. L. B. Bailey, in L. B. Bailey, ed., *Folate in Health and Disease*, Marcel Dekker Inc., New York, 1995, p. 123.
14. K. H. Bassler, E. Gröhn, D. Loew, and K. Pietrzik, *Vitamin-Lexikon*, G. Fischer Verlag, New York, 1992, p. 127.
15. J. Leichter, A. F. Landymore, and C. L. Krumdieck, *Am. J. Clin. Nutr.* **32**, 92 (1979).
16. S. P. Rothenberg, M. P. Iqbal, and M. Da Costa, *Anal. Biochem.* **103**, 152 (1980).
17. J. F. Gregory, *Adv. Food Nutr. Res.* **33**, 1 (1989).
18. J. F. Gregory, *Food Technol.* **42**, 230 (1988).
19. H. E. Sauberlich, M. J. Kretsch, J. H. Skala, H. L. Johnson, and P. C. Taylor, *Am. J. Clin. Nutr.* **46**, 1016 (1987); J. F. Gregory, in L. B. Bailey, ed., *Folate in Health and Disease*, Marcel Dekker Inc., 1995, p. 195.
20. IUPAC-IUB Joint commission on biochemical nomenclature (JCBN), *Eur. J. Biochem.* **168**, 251 (1987); *Pure Appl. Chem.* **59**, 834 (1987).
21. D. Bhatia, ed., in *Encyclopedia of Food Science and Technology*, John Wiley & Sons, Inc., New York, 1991, p. 2770.
22. C. W. Waller and co-workers, *J. Am. Chem. Soc.* **70**, 19 (1948).
23. F. Weygand and V. Schmied-Kowarzik, *Chem. Ber.* **82**, 333 (1949).
24. S. Uyeo, S. Mizukami, T. Kubota, and S. Takagi, *J. Am. Chem. Soc.* **72**, 5339 (1950).
25. **Rus. Pat. 7405995** (1974), K. Ito, H. Fukushima, and K. Nakagawa (to Nisshin Flour Milling Co., Ltd.).
26. W. Traube, *Berliner Berichte.* **33**, 1371 (1900).
27. **Ger. Pat. 3403468 A1** (Aug. 8, 1985), H. Blum and G. Dreesmann (to Federal Republic of Germany).
28. **Eur. Pat. 444266 A1** (Sept. 4, 1991), A. Hunds and W. Rogler (to Huls AG).
29. **Rus. Pat. 59204158** (1974) (to Daicel Chem Ind., KK).
30. **Ger. Pat. 3419817 A1** (Dec. 6, 1984), K. Yoshida, T. Niinobe, and T. Baba (to Takeda Chemical Industries Ltd.).
31. **Ger. Pat. 3605484A** (Aug. 8, 1987), W. Deinhammer and H. Petersen (to Walker-Chemie GmbH); **Eur. Pat. 394968 A** (Oct. 31, 1990), B. D.

Dombek and T. T. Wenzel (to Union Carbide Chemical).

32. J. R. Piper, G. S. McCaleb, and J. A. Montgomery, *J. Heterocycl. Chem.* **24**, 279 (1987).

33. C. M. Baugh and E. Shaw, *J. Org. Chem.* **29**, 3610 (1964).

34. **Ger. Pat. 2741383** (Feb. 22, 1979), E. Catalucci (to Lonza AG).

35. J. R. Piper and J. A. Montgomery, *J. Org. Chem.* **42**, 208 (1977).

36. J. R. Piper and J. A. Montgomery, *J. Heterocycl. Chem.* **11**, 279 (1974); **U.S. Pat. 4,077,957A** (Mar. 7, 1978), J. A. Montgomery and J. R. Piper (to U.S. Sec. Dept. Health); **U.S. Pat. 4,079,056** (Mar. 13, 1978), J. A. Montgomery and J. R. Piper (to U.S. Sec. Dept. Health).

37. J. H. Bieri and M. Viscontini, *Helv. Chim. Acta* **56**, 2905 (1973); K. Khalifa, P. K. Sengupta, J. H. Bieri, and M. Viscontini, *Helv. Chim. Acta* **59**, 242 (1976).

38. M. Sletzinger, D. Reinhold, J. Grier, M. Beachem, and M. Tishler, *J. Am. Chem. Soc.* **77**, 6365 (1955).

39. M. Viscontini and J. H. Bieri, *Helv. Chim. Acta* **54**, 2291 (1971).

40. **Swiss Pat. 255409** (1949) (to Hoffmann-La Roche Inc.).

41. **Ger. Pat. 3242193 A1** (May 17, 1984), F. Brunnmueller and M. Kroener (to BASF AG).

42. E. C. Taylor and T. Kobayashi, *J. Org. Chem.* **38**, 2817 (1973); E. C. Taylor, R. N. Henrie, and R. C. Portnoy, *J. Org. Chem.* **43**, 736 (1978).

43. B. Schirks, J. H. Bieri, and M. Viscontini, *Helv. Chim. Acta* **68**, 1639 (1985).

44. E. C. Taylor and K. Lenard, *Liebigs. Ann. Chem.* **726**, 100 (1969); E. C. Taylor and D. J. Dumas, *J. Org. Chem.* **46**, 1394 (1981).

45. **Ger. Pat. 3242195 A1** (May 17, 1984), F. Brunnmueller and M. Kroener (to BASF AG); **Eur. Pat. 175263 A2** (Mar. 26, 1986), H. Leninger, W. Littmann, J. Paust, and W. Trautmann (to BASF AG); **Eur. Pat. 175264 A2** (Mar. 26, 1986), H. Leininger, W. Littmann, and J. Paust (to BASF AG).

46. **Eur. Pat. Appl. 608 693 A2** (Aug. 3, 1994), C. Wehrli (to Hoffmann-La Roche AG).

47. J. F. Gregory and J. P. Toth, *J. Labelled Comp. Radiopharm.* **25**, 1349 (1988).

48. J. F. Gregory, *J. Agric. Food Chem.* **38**, 1073 (1990).

49. D. L. Hachey, L. Palladino, J. A. Blair, I. H. Rosenberg, and P. D. Klein, *J. Labelled Comp. Radiopharm.* **14**, 479 (1978).

50. J. F. Gregory and J. P. Toth, *Anal. Biochem.* **170**, 94 (1988).

51. P. A. Charlton, D. W. Young, B. Birdsall, J. Feeney, and G. C. K. Roberts, *J. Chem. Soc. Perkin Trans. I*, 1349 (1985).

52. C. M. Tatum, M. G. Fernald, and J. P. Schimel, *Anal. Biochem.* **103**, 255 (1980).

53. J. A. Blair and K. J. Saunders, *Anal. Biochem.* **34**, 376 (1970).

54. C. Zarow, A. M. Pellino, and P. V. Danenberg, *Prep. Biochem.* **12**, 381 (1983).

55. E. Khalifa, A. N. Ganguly, J. H. Bieri, and M. Viscontini, *Helv. Chim. Acta*, **63**, 2554 (1980).

56. D. B. Cosulich, J. M. Smith, and H. P. Broquist, *J. Am. Chem. Soc.* **74**, 4215 (1952).

57. J. E. Baggott and C. L. Krumdieck, *Biochemistry*, **18**, 1036 (1979).

58. C. M. Baugh, J. Stevens, and C. Krumdieck, *Biochim. Biophys. Acta*, **212**, 116 (1970).

59. B. L. Hutchings and co-workers, *J. Biol. Chem.* **170**, 323 (1947).

60. J. D. Cook, D. J. Cichowicz, S. George, A. Lawler, and B. Shane, *Biochemistry*, **26**, 530 (1987).

61. M. C. Li, R. Hertz, and D. M. Bergenstal, *New Engl. J. Med.* **259**, 66 (1958).

62. J. Burchanall and G. Hitchings, *Mol. Pharmacol.* **1**, 216 (1965).

63. G. M. Brown and H. Williamson, *Adv. Enzymol. Relat. Areas Mol. Biol.* **53**, 345 (1982).

64. G. M. Brown and H. Williamson, in F. C. Neidhart, ed., *Escherichia Coli and Salmonella Typhmurium*, Vol. 1, American Society for Microbiology, Washington, D.C., 1987, p. 521.

65. Y. Suzuki and G. M. Brown, *J. Biol. Chem.* **249**, 2405 (1974).

66. J. Kas and J. Cerna, *Methods Enzymol.* **66E**, 443 (1980).

67. P. M. Finglas, R. M. Faulks, and M. R. A. Morgan, *J. Micronutrient Anal.* **4**, 295 (1988).

68. C. L. Krumdieck, T. Tamura, and I. Eto, *Vitamins Horm.* **40**, 45 (1983).

69. G. Brubacher, W. Möller-Mulot, and D. A. T. Southgate, *Methods for Determination of Vitamins in Food*, Elsevier Applied Science Publisher, New York, 1985, p. 158.

70. D. B. McCormick and L. D. Wright, ed., *Methods Enzymol.* **66E**, 429 (1980).

71. H. Baker, S. P. Jaslow, and O. Frank, *J. Am. Geriatr. Soc.* **26**, 218 (1978).

72. S. A. Anderson and J. M. Talbot, *FDA Technical Report* FDA RF-82 13, Washington, D.C., 1981.

73. E. Giovannucci, E. B. Rimm, A. Ascherio, M. J. Stampfer, G. A. Colditz, and W. C. Willett, *J. Natl. Cancer Inst.* **87**, 265 (1995).

74. *MMWR Morb Mortal Wkly Rep.* **44**, 716 (1995).

75. A. Milunsky, J. Herschel, S. S. Jick, C. L. Bruell, D. S. Maclaughlin, K. J. Rothman, and W. Willett, *J. Am. Med. Assoc.* **262**, 2847 (1989).

76. C. Bower and J. F. Stanley, *Med. J. Austral.* **150**, 613 (1989).

77. P. S. Romano, N. J. Waitzman, R. M. Scheffler, and R. D. Pi, *Am. J. Public Health*, **85**, 667 (1995).

78. B. F. Satchell, *FDA Technical Report*, FDA 61RF-8781, Washington, D.C., 1996.

79. N. Whitehead, F. Reyner, and J. Lindenbaum, *J. Am. Med. Assoc.* **226**, 1421 (1973).

80. A. Bianchi, D. W. Chipman, A. Dreskin, and N. S. Rosensweig, *New Engl. J. Med.* **282**, 859 (1970).

81. K. S. McCully, *Am. J. Pathol.* **56**, 111 (1969).

82. C. S. Berwanger, J. Y. Jeremy, and G. Stansby, *Br. J. Surg.* **82**, 726 (1995).

83. J. Selhub, P. F. Jacques, A. G. Bostom, R. B. D'Agostino, P. W. Wilson, A. J. Belanger, D. H. O'Leary, P. A. Wolf, E. J. Schaefer, and I. H. Rosenberg, *N. Engl. J. Med.* **332**, 286 (1995).

84. J. L. Mills, J. M. McPartlin, P. N. Kirke, Y. J. Lee, M. R. Conley, D. G. Weir, and J. M. Scott, *Lancet*, **345**, 149 (1995).

85. A. F. Subar, G. Block, and L. D. James, *Am. J. Clin. Nutr.* **50**, 508 (1989).

86. *Nutrient Requirements of Poultry*, 9th. ed., National Research Council, National Academy Press, Washington, D.C., 1994; *Nutrient Requirements of Swine*, 9th ed., National Research Council, National Academy Press, Washington, D.C., 1988.

87. L. R. McDowell, *Vitamins in Animal Nutrition; Comparative Aspects to Human Nutrition*, Academic Press, Inc., San Diego, CA, 1989, pp. 298–325.

88. *Vitamin Nutrition for Poultry*, Dept. of Animal Health and Nutrition, Hoffmann-La Roche, Inc., Nutley, N.J., 1991.

89. *Vitamin Nutrition for Swine*, Dept. of Animal Health and Nutrition, Hoffmann-La Roche, Inc., Nutley, N.J., 1991.

90. J. Zimmerman, Z. Gihula, J. Selhub, and I. H. Rosenberg, *Int. J. Vit. Nutr. Res.* **59**, 151 (1989).

91. J. Zimmerman, J. Selhub, and I. H. Rosenberg, *Am. J. Clin. Nutr.* **46**, 518 (1987).

92. C. H. Young, F. M. Sirtnak, and M. Dembo, *J. Membrane Biol.* **79**, 285 (1984).

93. D. W. Horne, W. Y. Briggs, and C. Wagner, *J. Biol. Chem.* **253**, 3529 (1978).

94. G. B. Henderson, *Ann. Rev. Nutr.* **10**, 319 (1990).

95. C. Wagner, *Ann. Rev. Nutr.* **2**, 229 (1982).

96. C. Wagner, in L. B. Bailey, ed., *Folate in Health and Disease*, Marcel Dekker, Inc., New York, 1995, p. 23.

97. C. E. Butterworth and T. Tamura, *Am. J. Clin. Nutr.* **50**, 353 (1989).

98. R. Hunter, J. Barnes, H. F. Oakeley, and D. M. Matthews, *Lancet*, **1**, 61 (1970).

99. *New Engl. J. Med.* **237**, 713 (1947).

100. A. Hanck, in *Spektrum Vitamine*, **42**, 81 (1986).

101. A. F. Czeisal and J. Dudas, *New Engl. J. Med.* **327**, 1832 (1992); MRC Vitamin Study Research Group, *Lancet*, **338**, 131 (1991).

General References

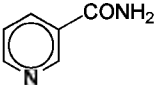
D. Bhatia, ed., in *Encyclopedia of Food Science and Technology*, John Wiley & Sons, Inc., New York, 1991, p. 2770.
J. F. Gregory, in L. B. Bailey, ed., *Folate in Health and Disease*, Marcel Dekker Inc., New York, 1995, p. 195.
T. Brody, in L. J. Machlin, ed., *Handbook of Vitamins*, Marcel Dekker Inc., New York, 1991, p. 453.
O. Isler and G. Brubacher, in O. Isler, G. Brubacher, S. Ghisla, and B. Kräutler, eds., *Vitamine II*, Georg Thieme Verlag, New York, 1982, p. 264.
S. K. Gaby and A. Bendich, in S. K. Gaby, A. Bendich, V. N. Singh, and L. J. Machlin, eds., *Vitamin Intake and Health*, Marcel Dekker Inc., New York, 1991, p. 175.

Thimma R. Rawalpally
Hoffmann-La Roche Inc.

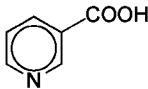


NIACIN, NICOTINAMIDE, AND NICOTINIC ACID

3-Pyridine carboxamide [98-92-0] (nicotinamide) (1) and 3-pyridine carboxylic acid [59-67-6] (nicotinic acid) (2) have a rich history and their early significance stems not from their importance as a vitamin but rather as products derived from the oxidation of nicotine. In 1867, Huber prepared nicotinic acid from the potassium dichromate oxidation of nicotine. Many years later, Engler prepared nicotinamide. Workers at the turn of the twentieth century isolated nicotinic acid from several natural sources. In 1894, Suzuki isolated nicotinic acid from rice bran, and in 1912 Funk isolated the same substance from yeast (1).



(1)



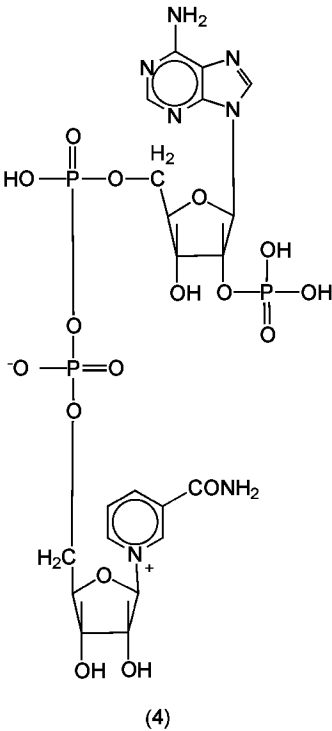
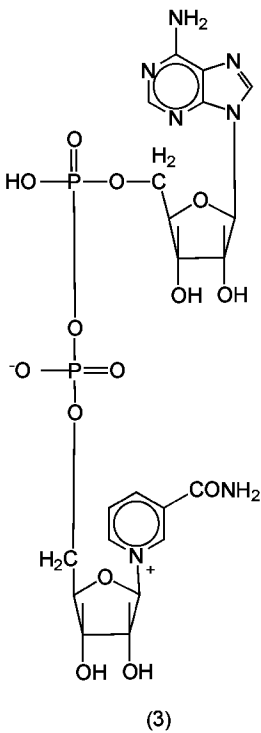
(2)

In 1913, Goldberger demonstrated that pellagra was due to a dietary deficiency. Pellagra had been earlier described by Thiery, who had coined the term *mal de la rosa* for this disease. Several decades later, Elvehjem and co-workers isolated nicotinamide from a liver extract and identified it as a pellagra-preventing factor (1).

There is considerable confusion regarding the nomenclature of these simple substances. The term niacin is a generic descriptor for pyridine 3-carboxylic acid and derivatives that exhibit the biological activity of nicotinic acid (1). However, niacin and vitamin PP (pellagra preventing) are frequently used interchangeably with nicotinic acid. Vitamin B₃ is often used as a designation for nicotinamide (1). The following is a list of common names for these substances.

Nicotinamide	Nicotinic acid
aminicotin	akotin
benicot	apelagrin
delonin amide	daskil
NAM	efacin
niacinamide	linic
niavit PP	niacin
nicasir	nicacid
nicosan 2	nicangin
nicovit	nicolar
pelmine	pellagrin
pelonin amide	pelonin
vitamin B	SR 4390
vitamin B ₃	

The biological importance of these compounds stems from their use as cofactors. Both nicotinamide and nicotinic acid are building blocks for coenzyme I (Co I), nicotinamide–adenine dinucleotide (NAD) (3) and coenzyme II (Co II), nicotinamide–adenine dinucleotide phosphate (NADP) (4) (2).



Chemical and Physical Properties

Nicotinamide is a colorless, crystalline solid. It is very soluble in water (1 g is soluble in 1 mL of water) and in 95° ethanol (1 g is soluble in 1.5 mL of solvent). The compound is soluble in butanol, amyl alcohol, ethylene glycol, acetone, and chloroform, but is only slightly soluble in ether or benzene. Physical properties are listed in Table 1.

Table 1. Physical Properties of Nicotinamide

Property	Value
molecular weight	122.12
melting point, °C	
stable modification	129–132
unstable modifications	
I	105
II	110
III	111
IV	113
V	116
boiling point, °C (0.067 Pa)	150–160
sublimation range, °C	80–100
true dissociation constants in water, at 20°C	
K_{b1}	2.24×10^{-11}
K_{b2}	3.16×10^{-14}
specific heat, kJ (kg·K) ^a	

solid, 55°C	1.30
65°C	1.34
75°C	1.39
liquid, 135°C	2.18
heat of solution in water, kJ kg ^a	148
heat of fusion, kJ kg ^a	381
density of melt, at 150°C, g cm ³	1.19

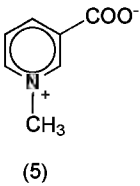
^a To connvert J to cal, multiply by 4.184.

Nicotinic acid is an amphoteric solid with needle-shaped crystals. It is less soluble than nicotinamide and its poor solubility in diethyl ether can be used as a basis to separate these compounds. Because nicotinamide has some solubility in ether, extraction of aqueous solutions of the acid and the amide with ether allows for selective extraction of the amide into the organic phase. Table 2 lists the physical properties of this vitamin.

Table 2. Physical Properties of Nicotinic Acid

Property	Value
molecular weight	123.11
melting point, °C	236–237
sublimation range	>150
density of crystals, g cm ³	1.473
dissociation constants in water, at 25°C	
<i>K</i> _a	1.50 × 10 ^{−5}
<i>K</i> _b	1.04 × 10 ^{−12}
isoelectric point in water, at 25°C, pH	3.42
pH of saturated aqueous solution	2.7
solubility, g L	
water	
0°C	8.6
38°C	24.7
100°C	97.6
ethanol, 96° o	
0°C	5.7
78°C	76.0
methanol	
0°C	63.0
62°C	345.0

The ring nitrogen of both the amide and the carboxylic acid can be quaternized and oxidized. *N*-methylnicotinate (trigonelline) (5) is an important component of green coffee beans and is converted to nicotinic acid during the roasting process. The reactivity of the 3-substituent parallels standard functional groups. Acid chlorides, esters, amides, and anhydrides have been prepared from nicotinic acid. Both the corresponding aldehyde and alcohol are available from the acid with a variety of reducing agents. Nicotinamide can be converted to nicotinic acid esters, the nitrile and acylamidines by routine methods.



Synthesis

Key intermediates in the industrial preparation of both nicotinamide and nicotinic acid are alkyl pyridines (Fig. 1). 2-Methyl-5-ethylpyridine (6) is prepared in a liquid-phase process from acetaldehyde. Also, a synthesis starting from ethylene has been reported. Alternatively, 3-methylpyridine (7) can be used as starting material for the synthesis of nicotinamide and nicotinic acid and it is derived industrially from acetaldehyde, formaldehyde (qv), and ammonia. Pyridine is the principal product from this route and 3-methylpyridine is obtained as a by-product. Despite this and largely due to the large amount of pyridine produced by this technology, the majority of the 3-methylpyridine feedstock is prepared in this fashion.

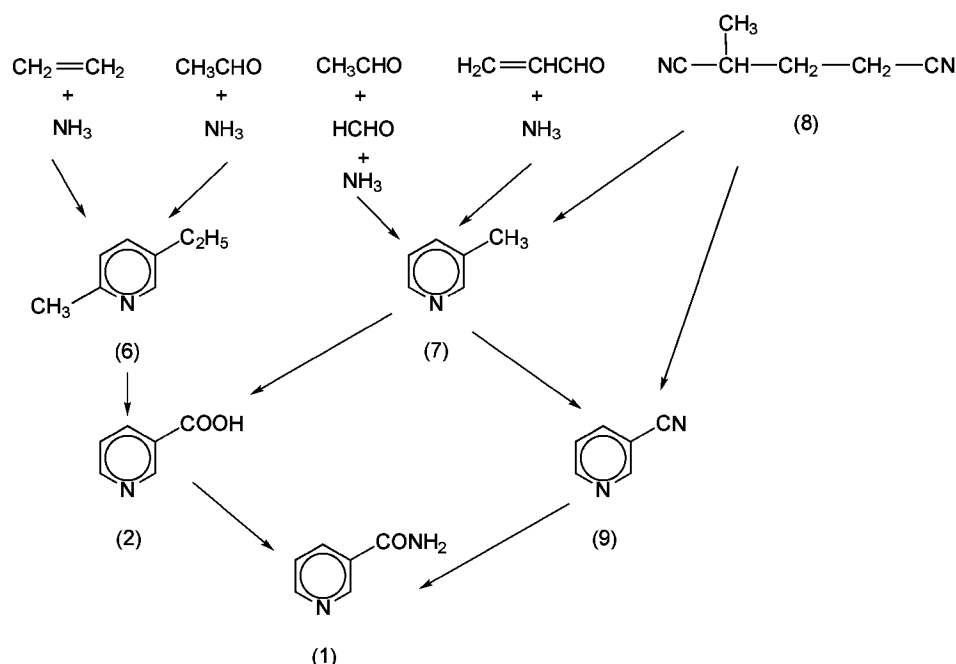


Fig. 1. Alkyl pyridines.

3-Methylpyridine can also be prepared by the condensation of acrolein and ammonia. Yields of 40–50% are obtained with pyridine as a by-product. Higher yields have been claimed when both propionaldehyde and acrolein have been used. A recent U.S. patent claims better selectivity if the cyclization is carried out in the presence of zeolites (3).

In an alternative approach, 2-methylglutaronitrile (8) is hydrogenated and cyclized to give high yields of 3-methylpyridine. The feedstock for this process is produced as a by-product of the production of adiponitrile. Oxidative cyclization of 2-methylglutaronitrile to nicotinonitrile (9) has been described (4).

The alkyl pyridines (6) and (7) can be transformed either to nicotinic acid or nicotinonitrile. In the case of nicotinic acid, these transformations can occur by either chemical or biological means. From an industrial standpoint, the majority of nicotinic acid is produced by the nitric acid oxidation of 2-methyl-5-ethylpyridine. Although not of industrial significance, the air oxidation has also been reported. Isocinchomeronic acid (10) (Fig. 2) is formed as an intermediate.

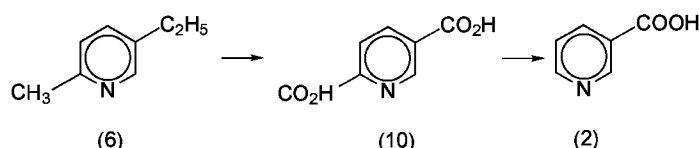


Fig. 2. Formation of isocinchomeronic acid (10).

Although an inherently more efficient process, the direct chemical oxidation of 3-methylpyridine does not have the same commercial significance as the oxidation of 2-methyl-5-ethylpyridine. Liquid-phase oxidation procedures are typically used (5). A Japanese patent describes a procedure that uses no solvent and avoids the use of acetic acid (6). In this procedure, 3-methylpyridine is combined with cobalt acetate, manganese acetate and aqueous hydrobromic acid in an autoclave. The mixture is pressurized to 101.3 kPa (100 atm) with air and allowed to react at 210°C. At a 32% conversion of the picoline, 19% of the acid was obtained. Electrochemical methods have also been described (7).

In more recent work, several groups have reported on fermentative approaches to nicotinic acids from 3-picoline. *Rhodococcus* (8), *Acinetobacter* (9), and *Pseudomonas* (10) have found utility in this application.

Nicotinonitrile is produced by ammoxidation of alkylpyridines (11–24). A wide variety of different catalysts have been developed for this application. For example, a recent patent describes a process in which 3-methylpyridine is reacted over a molybdenum catalyst supported on silica gel. The catalyst ($PV_4Mo_{12}O_{36}$) was prepared from NH_4VO_3 , H_3PO_4 , and $(NH_4)_5Mo_7O_{24}$. Reaction at 380°C at a residence time of 2.5 seconds gave 95% of nicotinonitrile at a 99% conversion (16).

Conversion of the nitrile to the amide has been achieved by both chemical and biological means. Several patents have described the use of modified Raney nickel catalysts in this application (25,26). Also, alkali metal perborates have demonstrated their utility (27). Typically, the hydrolysis is conducted in the presence of sodium hydroxide (28–31). Owing to the fact that the rate of hydrolysis of the nitrile to the amide is fast as compared to the hydrolysis of the amide to the acid, good yields of the amide are obtained. Other catalysts such as magnesium oxide (32), ammonia (28,29,33), and manganese dioxide (34) have also been employed.

Since the late 1980s, there has been considerable activity in the development of fermentative approaches for the preparation of the amide from the nitrile. Organisms such as *Achromobacter* (35), *Agrobacterium* (36), *Streptomyces* (37), *Rhodococcus* (38,39), and *Corynebacterium* (40) have been described. Purified enzymes in either free (41) or immobilized form (42,43) have also been used in this application.

Nicotinic acid has also been produced by microbial means from the nitrile. *Rhodococcus* (44,45) has frequently been used in this regard. Interestingly, irradiation of a *Corynebacterium* suspension during the fermentation led to higher yields of nicotinic acid (46).

Biosynthesis

A significant pathway for the formation of nicotinic acid mononucleotides begins with tryptophan (11). In the first step, cleavage of the indole ring is catalyzed by the enzyme L-tryptophan-2,3-dioxygenase. This step is rate determining unless it is hormonally induced by glucocorticoids and even tryptophan itself. The resulting *N*-formylkynurenine (12) is deformylated to the free aniline derivative, kynurenine (13). Hydroxylation to yield (14) is followed by oxidative removal of alanine to form 3-hydroxyanthranilic acid (15). Oxidative cleavage of the aromatic ring to the semialdehyde (16) is followed by dehydration to the important metabolite, quinolinic acid (17). Quinolinic acid is decarboxylated with concomitant alkylation at nitrogen to yield nicotinic acid mononucleotide (18) (Fig. 3).

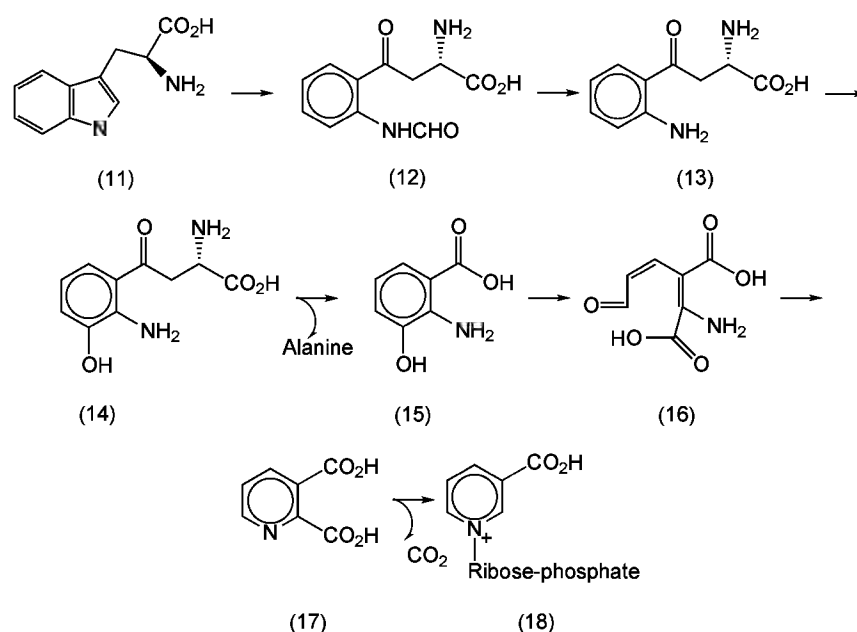


Fig. 3. Formation of nicotinic acid mononucleotides.

The result of this biosynthesis is that the product is nicotinic acid mononucleotide rather than free nicotinic acid. Ingested nicotinic acid is converted to nicotinic acid mononucleotide which, in turn, is converted to nicotinic acid adenine dinucleotide. Nicotinic acid adenine dinucleotide is then converted to nicotinamide adenine dinucleotide. If excess nicotinic acid is ingested, it is metabolized into a series of detoxification products (Fig. 4). Physiological metabolites include *N*-methylnicotinamide (19) and *N*-methyl-6-pyridone-2-carboxamide (24) (1).

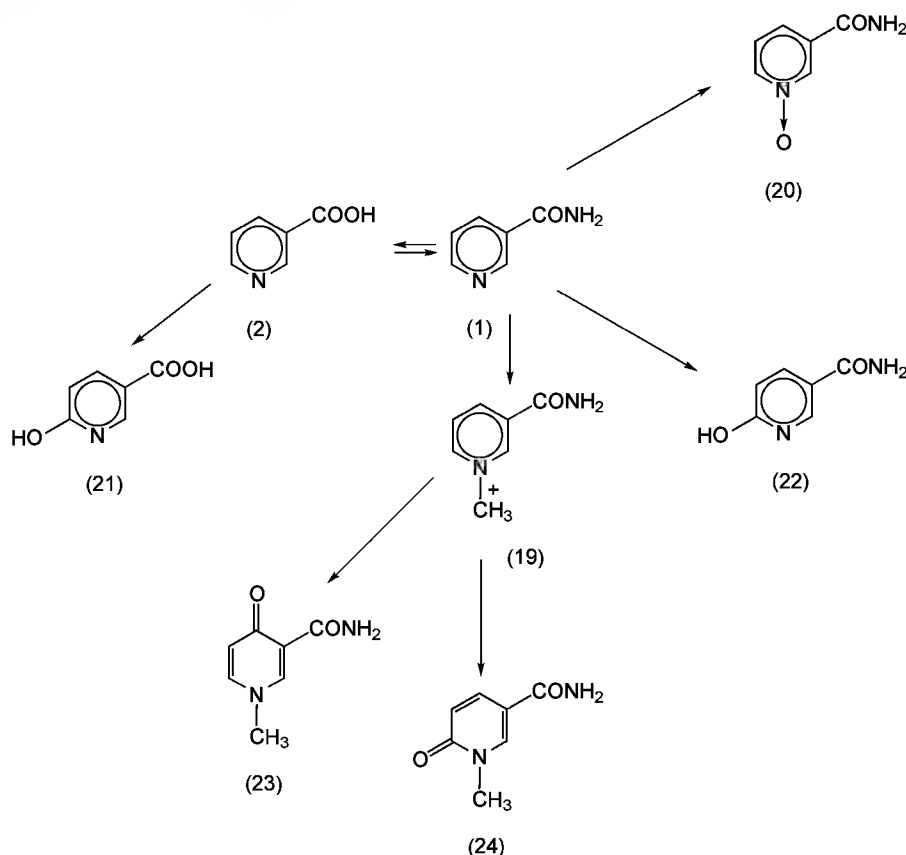


Fig. 4. Detoxification products as the result of digestion of excess nicotinic acid.

Nicotinamide is incorporated into NAD and nicotinamide is the primary circulating form of the vitamin. NAD has two degradative routes: by pyrophosphatase to form AMP and nicotinamide mononucleotide and by hydrolysis to yield nicotinamide adenosine diphosphate ribose.

Analytical Methods

Both nicotinic acid and nicotinamide have been assayed by chemical and biological methods. Owing to the fact that niacin is found in many different forms in nature, it is important to indicate the specific analyte in question. For example, if biological assay procedures are used, it is necessary to indicate whether the analysis is to determine the quantity of nicotinic acid or if niacin activity is the desired result of the analysis. If nicotinic acid is desired, then a method specific for nicotinic acid should be used. If quantitation of niacin activity is the desired outcome, then all compounds (bound and unbound) which behave like niacin will assay biologically for this substance (1).

The Kunig reaction (Fig. 5) has been used to determine the amount of nicotinic acid and nicotinamide. In this procedure, quaternization of the pyridine nucleus by cyanogen bromide is followed by ring opening to generate the putative dialdehyde intermediate. Reaction of this compound with an appropriate base, such as *p*-methylaminophenol sulfate (47) or sulfanilic acid (48), generates a dye. The concentration of this dye is determined colorimetrically.

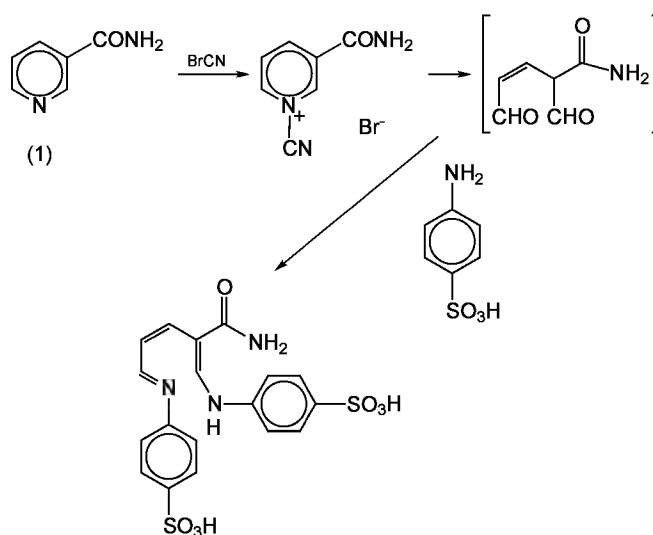


Fig. 5. The Künig reaction.

In the case of nicotinamide, the color yield is often low. This problem can be circumvented by either hydrolysis to nicotinic acid or by conversion of the amide to a fluorescent compound. Treatment of nicotinamide with methyl iodide yields the quaternary ammonium salt, *N*-methyl nicotinamide (5). Reaction of this compound with acetophenone yields a fluorescent adduct (49). Other carbonyl compounds have also been used (50–54).

For more specific analysis, chromatographic methods have been developed. Using reverse-phase columns and uv detection, hplc methods have been applied to the analysis of nicotinic acid and nicotinamide in biological fluids such as blood and urine and in foods such as coffee and meat. Derivatization techniques have also been employed to improve sensitivity (55). For example, the reaction of nicotinic amide with DCCI (*N*'-dicyclohexyl-*O*-methoxycoumarin-4-yl)methyl isourea to yield the fluorescent coumarin ester has been reported (56). After separation on a reversed-phase column, detection limits of 10 pmol for nicotinic acid have been reported (57).

Owing to poor volatility, derivatization of nicotinic acid and nicotinamide are important techniques in the gc analysis of these substances. For example, a gc procedure has been reported for nicotinamide using a flame ionization detector at detection limits of $\sim 0.2 \mu\text{g}$ (58). The nonvolatile amide was converted to the nitrile by reaction with heptafluorobutyric anhydride (56). For a related molecule, quinolinic acid, fmol detection limits were claimed for a gc procedure using either packed or capillary columns after derivatization to its hexafluoroisopropyl ester (58).

As with many of the vitamins, biological assays have an important historical role and are widely used. For example, microbiological assays use *Lactobacillus plantarum* ATCC No. 8014 (57,59) or *Lactobacillus arabinosus* (60). These methods are appropriate for both nicotinamide and nicotinic acid. Selective detection of nicotinic acid is possible if *Leuconostoc mesenteroides* ATCC No. 9135 is used as the test organism (61). The use of microbiological assays have been reviewed (62).

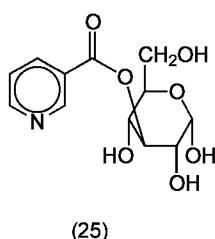
Bioassays procedures have been developed in species such as chicks which have been fed a niacin-deficient diet. Due to the fact that, for example, tryptophan is a biological precursor of niacin, niacin can be produced from other sources (55). As a result, the tryptophan content of the diet has to be monitored carefully for accurate results.

Specifications for niacin and nicotinamide for food use are given in the *Food Chemicals Codex* (63) and for pharmaceutical use in the *United States Pharmacopeia* (64). The *Codex* also gives specifications for nicotinamide ascorbate.

Occurrence

Nicotinamide and nicotinic acid occur in nature almost exclusively in the bound form. In plants, nicotinic acid is prevalent whereas in animals nicotinamide is the predominant form. This nicotinamide is exclusively in the form of NAD and NADP.

Nicotinic acid is found in plants associated with both peptides and polysaccharides. For example in wheat bran, two forms are described: a peptide with a molecular weight of approximately 12,000 and a carbohydrate complex that is called niacytin. Polysaccharides isolated from wheat bran have been found to contain 1.05% nicotinic acid in bound form. Hydrolysis yielded a fragment identified as β -3-*O*-nicotinoyl-D-glucose (25).



From a bioavailability standpoint, the fact that a significant amount of nicotinic acid is in a bound form has important biological consequences. Poor bioavailability stems from the fact that the ester linkage is resistance to digestive enzymes. In the case of corn, this condition can be alleviated if corn is pretreated with alkali. This food preparation method is frequently practiced in Mexico for the preparation of tortillas.

Nicotinamide and nicotinic acid are prevalent in many common foodstuffs and are especially concentrated in brewer's yeast, wheat germ and liver. In this regard, tryptophan is considered a provitamin and is assigned a niacin equivalent of 1/60. The following lists the vitamin B₃ content of many common foodstuffs and in Table 3, values of vitamin B₃ content are compared to niacin potential from tryptophan.

corn, yellow
corn, gluten feed
rye
oats
milo
barley
wheat
wheat bran
tapioca flour
potatoes

Fig. 6. Coenzymes required for dehydrogenase.

In addition to their *in vivo* function, these cofactors have important *in vitro* functions as co-factors in enzymatic synthesis. Because these co-factors are too expensive to be used in equimolar amounts, many methods have been developed for their regeneration. These include chemical (66), biological (67) and electrochemical (68) methods. Enzymatic regeneration has found particular utility in this application (69).

Deficiency

A deficiency of niacin manifests itself in the disease pellagra. The symptoms of pellagra include dermatosis, dementia, and diarrhea. The dermatosis is expressed as lesions. These lesions are most frequently found on the hands, wrists, elbows, face and neck, knees, and the feet. At the onset of the disease, the affected areas resemble sunburn, progressing to keratosis and scaling of pigmented skin. Other symptoms include insomnia, loss of appetite, weight and strength loss, soreness of the tongue and mouth, indigestion, diarrhea, abdominal pain, burning sensations, vertigo, headache, numbness, nervousness, distractability, apprehension, mental confusion, and forgetfulness (1).

A deficiency of niacin also affects the nervous system. Numbness is initially observed and later, paralysis, particularly in the extremities is common. Severe cases are characterized by tremor and a spastic or ataxic gait and are frequently associated with peripheral neuritis. Left untreated, severe thought disorders can ensue (1).

Requirements

The RDA for niacin is based on the concept that niacin coenzymes participate in respiratory enzyme function and 6.6 niacin equivalents (NE) are needed per intake of 239 kJ (1000 kcal). One NE is equivalent to 1 mg of niacin. Signs of niacin deficiency have been observed when less than 4.9 NE 239 kJ or less than 8.8 NE per day were consumed. Dietary tryptophan is a rich source of niacin and the average diet in the United States contains 500–1000 mg of tryptophan. In addition, the average diet contains approximately 8–17 mg of niacin. In total, these two quantities total 16–34 NE daily. Table 5 lists the RDA and U.S. RDA for niacin (69).

Table 5. RDA and U.S. RDA for Niacin

Category	Niacin, mg NE
<i>RDA</i>	
infants	
0.0–0.5 yr	5
0.5–1.0 yr	6
children	
1–3 yr	9
4–6 yr	12
7–10 yr	13
males	
11–14 yr	17
15–18 yr	20
19–50 yr	19
>51 yr	15
females	
11–50 yr	15
>51 yr	13
pregnant	17
lactating	20
<i>U.S. RDA</i>	
infants and children <4 yr	9
adults and children >4 yr	20
pregnant or lactating women	20

Safety

Despite structural similarities, the pharmacological consequences of excesses of these substances are quite different. Due to the interest in the effects of nicotinic acid on atherosclerosis, and in particular its use based on its ability to lower serum cholesterol, the toxicity of large doses of nicotinic acid has been evaluated. For example, in a study designed to assess its ability to lower serum cholesterol, only 28% of the patients remained in the study after receiving a large initial dose of 4 g of nicotinic acid due to intolerance at these large doses (70).

Nicotinamide can also be toxic to cells at concentrations that increase the NAD levels above normal. Individuals consuming nicotinamide at levels of 3 g d for extended periods (3–36 months) have experienced various side effects such as heartburn, nausea, headaches, hives, fatigue, sore throat, dry hair, and tautness of the face (1).

Uses

Both nicotinic acid and nicotinamide have been used in the enrichment of bread, flour, and other grain-derived products. Animal feed is routinely supplemented with nicotinic acid and nicotinamide. Nicotinamide is also used in multivitamin preparations. Nicotinic acid is rarely used in this application. The amide and carboxylic acid have been used as a brightener in electroplating baths and as stabilizer for pigmentation in cured meats.

Derivatives

Nicotiny alcohol (3-pyridinylcarbinol, 3-pyridinemethanol) (27) has use as an antilipemic and peripheral vasodilator. It is available from either the reductions of nicotinic acid esters or preferably, the reduction of the nitrile to the amine followed by diazotation and nucleophilic displacement. It is frequently administered in the form of the tartrate (Fig. 7). Nicotinic acid is frequently used as a salt in conjunction with basic drugs such as the peripheral vasodilator xanthinol niacinate (28). Nicotinic acid and its derivatives have widespread use as antihyperlipidemic agents and peripheral vasodilators (1).

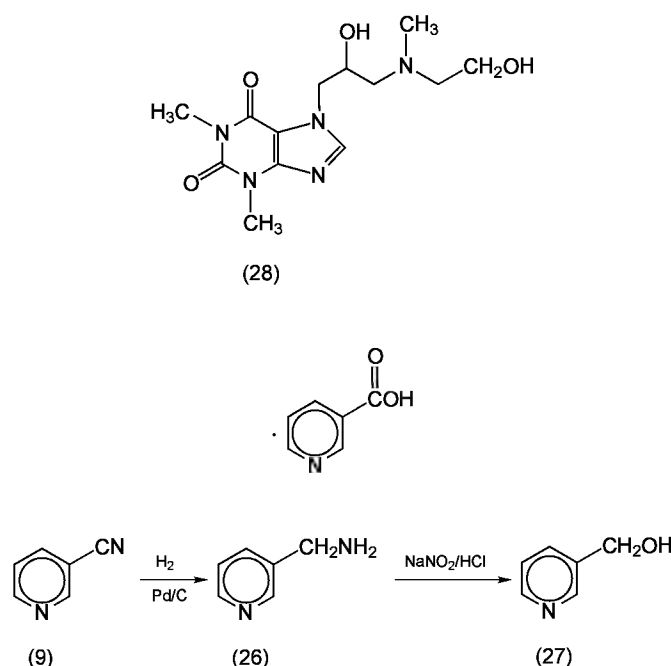


Fig. 7. Formation of nicotiny alcohol (27).

Economic Aspects

U.S. manufacturers of niacin and niacinamide include Nepera, Inc. and Reilly Industries, Inc. U.S. suppliers include BASF Corporation, Hoffmann-La Roche Inc., and Rhône-Poulenc. Western European producers and suppliers include Degussa, Rhône-Poulenc, BASF, Hoffmann-La Roche, and Lonza (71). In 1995, the prices for niacin and nicotinamide were \$9.75/kg and \$9.25/kg, respectively (72,73).

BIBLIOGRAPHY

"Nicotinic Acid and Nicotinamide" in *ECT* 1st ed., Vol. 9, pp. 305–313, by A. P. Sachs and J. F. Couch, U.S. Dept. of Agriculture; "Nicotinic Acid" under "Vitamins" in *ECT* 2nd ed., Vol. 21, pp. 509–542, by C. del Río-Estrada, Universidad Nacional de Mexico, and H. W. Dougherty, Merck Sharp & Dohme; "Nicotinamide and Nicotinic Acid" under "Vitamins" in *ECT* 3rd ed., Vol. 24, pp. 59–93, by H. Offermanns and co-workers, Degussa Co.

- J. van Eys, in L. Machlin, ed., *Handbook of Vitamins*, Marcel Dekker, New York, 1991, pp. 311–340.
- D. Bhatia, ed., *Encyclopedia of Food Science and Technology*, John Wiley & Sons, Inc., New York, 1991, pp. 2276–2783.
- U.S. Pat. 5,395,940 (Mar. 7, 1995), P. J. Angevine, C. T. Chu, and T. C. Potter (to Mobil Oil Corp.).
- U.S. Pat. 4,721,789 (Dec. 6, 1988), R. Dicosimo, J. D. Burrington, and R. K. Grasselli (to Standard Oil Co.).
- Jpn. Pat. 94-26603 (Feb. 24, 1994), Y. Asamidori, I. Hashiba, and S. Takigawa (to Nissan Chemical Ind Ltd, Japan).
- World Pat. 9305022 A1 (Mar. 18, 1993), M. Hatanaka and N. Tanaka (to Nissan Chemical Industries, Ltd., Japan).
- U.S. Pat. 4,482,439 A (Nov. 13, 1984), J. E. Toomey, Jr. (to Reilly Tar and Chemical Corp.).
- Jpn. Pat. 93-193212 (July 8, 1993), T. Ishikawa, K. Maeda, Y. Mukohara, and T. Kojima (to Nippon Soda Co., Japan).
- Jpn. Pat. 90-80698 (Mar. 30, 1990), F. C. Fujita (to Nitto Chemical Industry Co., Ltd.).
- Eur. Pat. 442430 A2 (Aug. 21, 1991), A. Kiener (to Lonza, AG).
- Eur. Pat. 684074 A1 (Nov. 29, 1995), Y. K. Lee, C. H. Shin, T. S. Chang, D. H. Cho, and D. K. Lee (to Korea Research Institute of Chemical Technology, South Korea).
- Eur. Pat. 525367 A1 (Feb. 3, 1993), Y. Onda, K. Tsukahara, and N. Takahashi (to Mitsubishi Gas Chemical Co.).
- U.S. Pat. 5,028,713 (July 2, 1991), R. Dicosimo, J. D. Burrington, and R. K. Grasselli (to Standard Oil Co.).
- U.S. Pat. 4,876,348 (Oct. 24, 1989), R. Dicosimo, J. D. Burrington, and D. D. Suresh (to Standard Oil Co.).
- Eur. Pat. 339680 A2 (Nov. 2, 1989), M. Saito, K. Tsukahara, K. Yamada, and H. Imai (to Mitsubishi Gas Chemical Co., Inc.).
- Int. Pat. 163169A (Aug. 20, 1988), B. V. Suvorov and co-workers (to the Institute of Chemical Sciences, Academy of Sciences, Kazakh SSR).
- Ger. Pat. 241903 A1 (Jan. 7, 1987), B. Luecke and co-workers (to Akademie der Wissenschaften der DDR).
- Eur. Pat. 253360 A2 (Jan. 20, 1988), S. Shimizu and co-workers (to Koei Chemical Co., Ltd., Japan).
- Rus. Pat. 86-154741 (July 1, 1986), N. Abe, M. Dojo, and A. Iguchi (to Koei Chemical Industry Co., Ltd.).
- Jpn. Pat. 85-169554 (July 30, 1985), H. Sato and K. Hirose (to Sumitomo Chemical Co., Ltd., Japan).
- U.S. Pat. 4,603,207 A (July 29, 1986), R. Dicosimo, J. D. Burrington, and D. D. Suresh (to Standard Oil Co., Ohio).
- Eur. Pat. 96-100676 (Jan. 18, 1996), L. V. Hippel, A. Neher, and D. Arntz (to Degussa Aktiengesellschaft, Germany).
- World Pat. 9532055 (Nov. 30, 1995), D. C. Sembaer, F. A. Ivanovskaya, and E. M. Guseinov (to Lonza Ltd, Switzerland).
- World Pat. 9532054 (Nov. 30, 1995), B. V. Suvorov, L. A. Stepanova, B. N. Anatoljevna, J. R. Chuck, and D. Pianzola (to Lonza Ltd, Switzerland).
- Jpn. Pat. 93-206579 (Aug. 20, 1993), H. Hirayama (to Showa Denko KK, Japan).
- Eur. Pat. 85-306670 (Sept. 19, 1985), S. Asano and J. Kitagawa (to Mitsui Toatsu Chemicals, Inc.).
- World Pat. 9009988 A1 (Sept. 7, 1990), A. McKillop and D. Kemp (to Interlox Chemicals, Ltd.).
- U.S. Pat. 2,471,518 (May 31, 1949), B. F. Duesel and H. H. Friedman (to Pyridium Co.).
- U.S. Pat. 4,721,789 (Dec. 6, 1988), R. Dicosimo, J. D. Burrington, and R. K. Grasselli (to Standard Oil Co.).
- Brit. Pat. Appl. 11357/55 (Aug. 28, 1942), B. F. Duesel and H. H. Friedman (to Pyridium Co.).
- Ger. Offen. Appl. 2,517,054 (Apr. 17, 1975), H. Beschke, H. Friedrich, K.-P. Muller, and G. Schreyer (to Degussa Co.).
- C. B. Rosas and G. B. Smith, *Chem. Eng. Sci.* **35**, 330 (1975).
- Brit. Pat. 777,517 (Mar. 28, 1956), E. J. Gasson and D. J. Hadley (to The Distillers Co., Ltd.).
- U.S. Pat. 4,008,241 (Feb. 15, 1977), A. P. Gelbein and co-workers (to the Lummus Co.).
- Ger. Pat. 2,1313,8134 (Oct. 19, 1972), L. R. Haeefe (to R. J. Reynolds Tobacco Co.).
- Jpn. Pat. 93-18645 (Feb. 5, 1993), A. Hatamori, M. Shimada, and K. Ishiwatari (to Mitsui Toatsu Chemicals).
- Eur. Pat. 568072Y A2 (Feb. 27, 1991), Y. Takashima, K. Kumagai, and S. Mitsuda (to Sumitomo Chemicals Co.).

38.

Jpn. Pat. 90-80697 (Mar. 30, 1990), F. To and C. Fujita (to Nitto Chemical Industry Co., Ltd.).

39.

Jpn. Pat. 89-81137 (Mar. 31, 1989), H. Yamada, T. Nagasawa, and H. Shimizu (to Nitto Chemical Industry Co., Ltd.).

40.

Eur. Pat. 204555 A2 (Dec. 10, 1986), K. Kawakami, T. Tanabe, and O. Nagano (to Asahi Chemical Industry Co., Ltd.).

41.

Jpn. Pat. 82-91408 (May 31, 1982) (to Nitto Chemical Industry Co., Ltd.).

42.

Eur. Pat. 579907 A1 (Jan. 26, 1994), K. Yamada, M. Ochiai, Y. Yotsumoto, Y. Morimoto, and Y. Teranishi (to Mitsubishi Kawei Corp., Japan).

43.

J. Eyal and M. Charles, *Ann, N.Y. Acad. Sci.* **589**, 678 (1990).

44.

Jpn. Pat. 86-109401 (May 15, 1986), K. Kawakami and T. Tanabe (to Technology Research Association for New Application Development for Light-Weight Fractions, Japan).

45.

Eur. Pat. 91-102938 (Feb. 27, 1991), H. Yamada, T. Nagasawa, and T. Nakamura (to Nitto Chemical Co., Ltd., Japan).

46.

C. D. Mathew, T. Nagasawa, and M. Kobayashi, *Appl. Environ. Microbiol.* **54**, 1030–1032 (1988).

47.

Eur. Pat. 86-100305 (Jan. 11, 1986), K. Enomoto, Y. Sato, Y. Nakashima, A. Fujiwara, and T. Doi (to Nitto Chemical Industry Co., Ltd., Japan, Mitsubishi Rayon Co., Ltd.).

48.

F. E. Friedemann and E. J. Frazier, *Arch. Biochem.* **26**, 361 (1950).

49.

Official Methods of Analysis, 8th ed., W. Horwitz, ed., Association of Official Agricultural Chemists, Washington, D.C., 1955, p. 826.

50.

B. R. Clark, in D. B. McCormick and L. D. Wright, eds., *Methods in Enzymology*, Vol. 66, Academic Press, Inc., New York, 1980, p. 5.

51.

J. W. Huff and W. A. Perlzweig, *J. Biol. Chem.* **167**, 157 (1947).

52.

N. Levitas, J. Robinson, F. Rosen, J. W. Huff, and W. A. Perlzweig, *J. Biol. Chem.* **167**, 169 (1947).

53.

H. B. Burdch and co-workers, *J. Biol. Chem.* **212**, 897 (1955).

54.

G. H. Carpenter and E. Kodicek, *Biochem. J.* **46**, 421 (1950).

55.

G. A. Goldsmith and O. N. Miller, in P. Gyurgy and W. N. Pearson, eds., *The Vitamins*, 2nd ed., Vol. VII, Academic Press, Inc., New York, 1967, p. 137.

56.

Y. Tsuruta, K. Kohashi, S. Ishida, and Y. Ohkura, *J. Chromatogr.* **309**, 309 (1984).

57.

K. Shibata and T. Shimono, in A. D. P. De Leenheer, W. E. Lambert, and H. J. Nelis, eds., *Modern Chromatographic Analysis of Vitamins*, 2nd ed., Marcel Dekker, Inc., New York, 1992, pp. 285–317.

58.

A. Tanaka, M. Iijima, Y. Kikuchi, Y. Hoshino, and N. Nose, *J. Chromatogr.* **466**, 307 (1989).

59.

M. P. Heyes and S. P. Markey, *Anal. Biochem.* **174**, 349 (1988).

60.

Official Methods of Analysis, Association of Official Analytical Chemists, Washington, D.C., 1980, pp. 759–761, 763–764.

61.

Methods of Vitamin Assay, Interscience Publishers, New York, 1966, pp. 172–176.

62.

B. C. Johnson, *J. Biol. Chem.* **159**, 227 (1945).

63.

M. R. S. Fox and G. M. Briggs, *J. Nutr.* **72**, 243 (1960).

64.

Food and Nutrition Board, National Research Council, *Food Chemicals Codex*, 3rd ed., National Academy Press, Washington, D.C., 1981, pp. 205–206.

65.

The United States Pharmacopeia XXII (USP XXII-NF XVII), United States Pharmacopeial Convention, Inc., Rockville, Md., 1990, pp. 943–944.

66.

H. Tagachi, M. Sakaguchi, and Y. Shimabayashi, *Agri. Bio. Chem.* **49**, 3467–3471 (1985).

67.

R. Ruppert, S. Hermann, and E. Steckhan, *J. Chem. Soc. Chem. Comm.*, 1150–1151 (1988).

68.

Y. Morita, H. Furui, H. Yoshii, and R. Matsuno, *J. Ferment. Bioeng.* **77**, 13–16 (1994).

69.

C. H. Wong and G. Whitesides, *Enzymes in Organic Chemistry*, Pergamon Press, Elsevier Science Ltd., Oxford, U.K., 1994.

70.

Committee on Dietary Allowances, Food and Nutrition Board, *Recommended Dietary Allowances*, 10th ed., National Academy of Sciences, Washington, D.C., 1989, pp. 137–142.

71.

H. K. Schock, in W. L. Holmes, L. H. Carlson, and R. Paoletts, eds., *Drugs Affecting Lipid Metabolism*, Plenum Press, New York, 1969, p. 405.

72.

Chemical Economics Handbook, SRI International, Menlo Park, Calif., July 1995.

73.

Chem. Mark. Rep. (Mar. 18, 1996).

General References

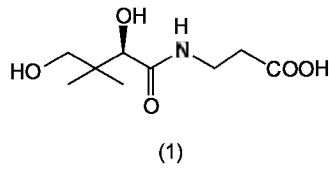
Refs. 2, 3, and 57 are good general references.

O. Isler, G. Brubacher, S. Ghisla, and B. K_utler, "Die Niacin-Gruppe (Vitamin PP)" in *Vitamine II*, Georg Thieme Verlag Stuttgart, New York, 1988, pp. 160–192.

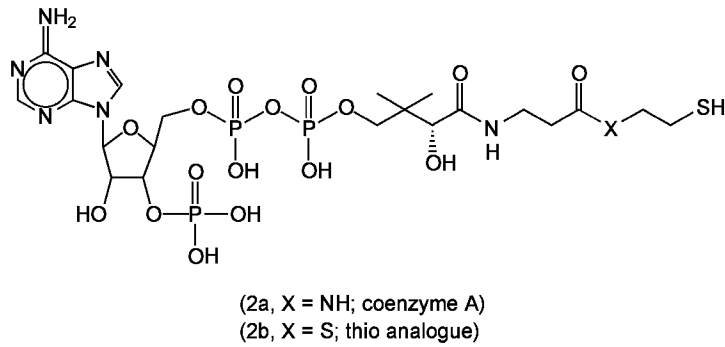
Susan D. Van Arnum
Hoffmann-La Roche Inc.

PANTOTHENIC ACID

(R)-Pantothenic acid [79-83-4] (1) is a member of the B-complex vitamins. It is a water-soluble vitamin and has been designated as vitamin B₅. It was discovered by Williams and co-workers in 1933 by extracting a growth factor for yeast from a wide range of biological tissues (1). Later the growth factor was identified as pantothenic acid (1). Independently two groups, Woolley and co-workers (2) in 1939 and Jukes (3) between 1939 and 1941, identified the liver filtrate factor in rats and antidermatitis factor in chicks. These factors later proved to be identical with pantothenic acid. The total synthesis of (R)-pantothenic acid was achieved in 1940 by three groups: Stiller and co-workers (4), Reichstein and Grössner (5), and Kuhn and Wieland (6).



After a full structural elucidation of coenzyme A in 1953 by Baddiley, it became evident that pantothenic acid is one of the components of coenzyme A (2) (7).



Biosynthesis of coenzyme A (CoA) in mammalian cells incorporates pantothenic acid. Coenzyme A, an acyl group carrier, is a cofactor for various enzymatic reactions and serves as either a hydrogen donor or an acceptor. Pantothenic acid is also a structural component of acyl carrier protein (ACP). ACP is an essential component of the fatty acid synthetase complex, and is therefore required for fatty acid synthesis. Free pantothenic acid is isolated from liver, and is a pale yellow, viscous, and hygroscopic oil.

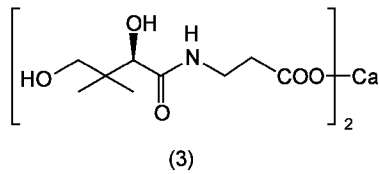
Occurrence, Source, and Bioavailability

Good food sources of pantothenic acid include yeast, chicken, beef, potatoes, oat cereal, vegetables, legumes, and whole grain. Pantothenic acid in foods and feedstuffs is fairly stable to ordinary means of cooking and storage, but appreciable losses have been reported as a result of canning and storage of some foods. Pantothenic acid is synthesized only by plants and microorganisms. Estimation of the dietary intake of pantothenic acid is difficult because the acid exists in the free form and is also incorporated in coenzyme A and fatty acid synthetase. Total pantothenic acid content present in the food is estimated after releasing the bound acid enzymatically (8,9).

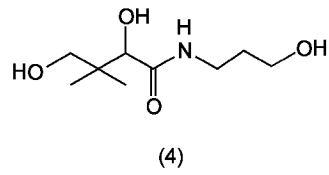
Relatively little is known about the bioavailability of pantothenic acid in human beings, and only approximately 50% of pantothenic acid present in the diet is actually absorbed (10). Liver, adrenal glands, kidneys, brain, and testes contain high concentrations of pantothenic acid. In healthy adults, the total amount of pantothenic acid present in whole blood is estimated to be 1 mg L. A significant (2–7 mg d) difference is observed among different age-group individuals with respect to pantothenic acid intake and urinary excretion, indicating differences in the rate of metabolism of pantothenic acid.

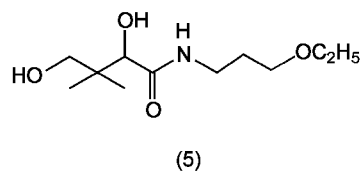
Physical and Chemical Properties

(R)-Pantothenic acid (1) contains two subunits, (R)-pantoic acid and β-alanine. The chemical abstract name is N-(2,4-dihydroxy-3,3-dimethyl-1-oxobutyl)-β-alanine (11). Only (R)-pantothenic acid is biologically active. Pantothenic acid is unstable under alkaline or acidic conditions, but is stable under neutral conditions. Pantothenic acid is extremely hygroscopic, and there are stability problems associated with the sodium salt of pantothenic acid. The major commercial source of this vitamin is thus the stable calcium salt (3) (calcium pantothenate).

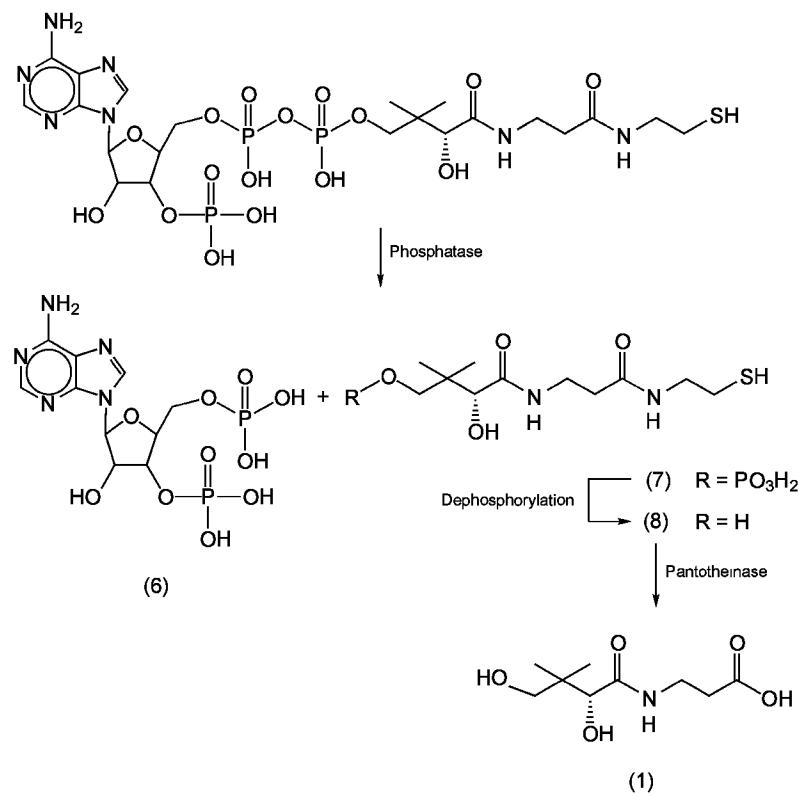


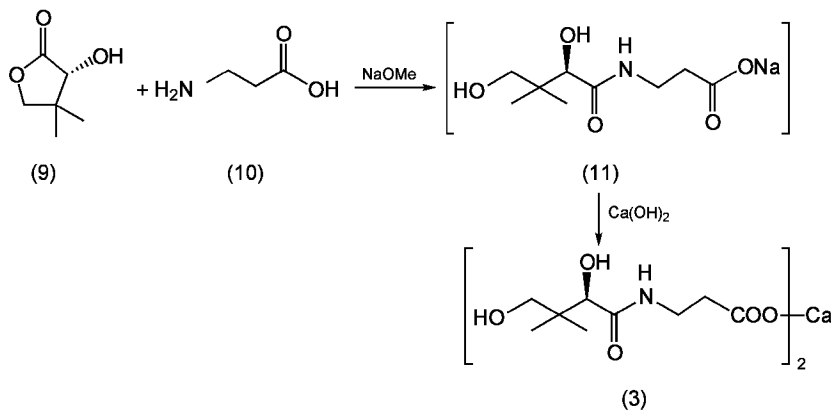
Panthenol (4) is the reduced form of pantothenic acid and is the pure form most commonly used. The alcohol is more easily absorbed and is converted into the acid *in vivo* (12). Both panthenol and pantyl ether are used in hair care products.



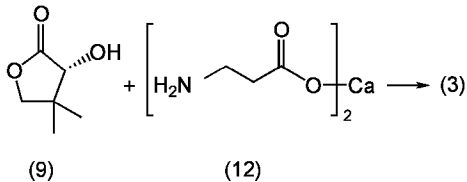


The biologically active R- or *d*-pantothenic acid can be obtained upon hydrolysis of coenzyme A with a combination of two enzymes, alkaline phosphatase and pantotheinase (13) (Fig. 1). The phosphatase catalyzes the selective cleavage of the phosphate bond in coenzyme A to afford adenosin-3'5'-diphosphate (6) and 4-phosphopantetheine (7). The latter substance is dephosphorylated enzymatically to yield pantetheine (8), which is rapidly converted by pantotheinase to pantothenic acid (1). Table 1 lists some physical properties of pantothenic acid and its derivatives.

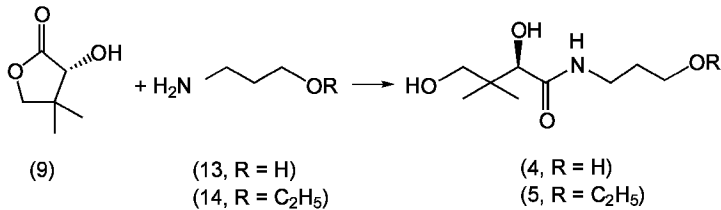




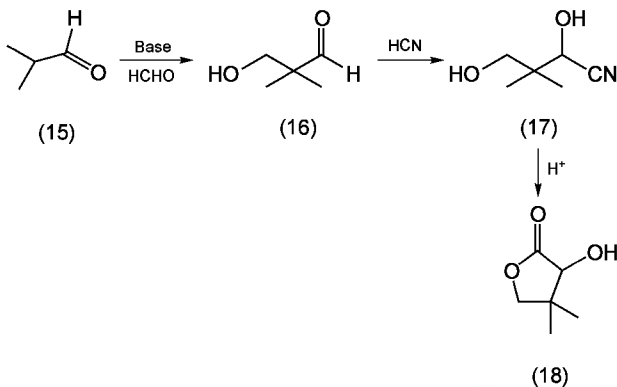
An alternative procedure for the preparation of (R)-calcium pantothenate (3) is to condense (R)-pantolactone (9) with the preformed calcium salt (12) of β -alanine (15).



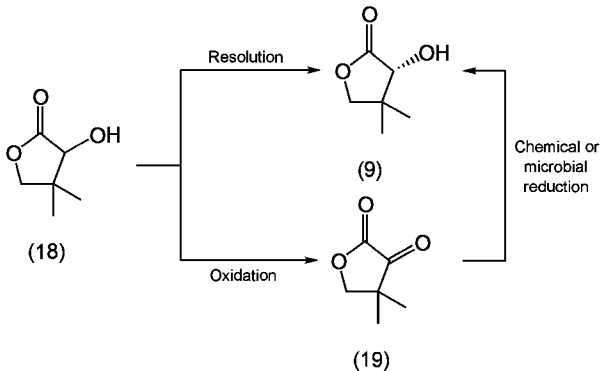
Similarly, panthenol (4) and pantyl ether (5) are prepared by condensing 3-aminopropanol (13) and 3-ethoxypropylamine (14) with (R)-pantolactone (16,17).



Racemic pantolactone is prepared easily by reacting isobutyraldehyde (15) with formaldehyde in the presence of a base to yield the intermediate hydroxyaldehyde (16). Hydrogen cyanide addition affords the hydroxy cyanohydrin (17). Acid-catalyzed hydrolysis and cyclization of the cyanohydrin (17) gives (R,S)-pantolactone (18) in 90% yield (18).

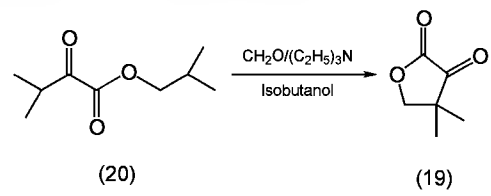


(R)-Pantolactone (9) is prepared in either of two ways: by resolution of the (R,S)-pantolactone mixture (18), or by stereoselective reduction of ketopantolactone (19) by chemical or microbial methods (19).



Chemical resolution of pantoic acid, readily available from (R,S)-pantolactone (18), is the most efficient method currently being practiced on an industrial scale. A wide variety of chiral resolving agents are being used, such as (+)-2-aminomethylpinane by BASF (20), 2-benzylamino-1-phenylethanol by Fuji (21), chiral aminoalcohols by Alps and Sumitomo (22,23), (1R)-3-endo-aminoborneol by Hoffmann-La Roche (24), and 1-ethylbenzylamine by Daicel (25). Other approaches include optical resolution by inclusion crystallization (26) and spontaneous resolution of (R,S)-pantolactone by seeding (27). (R,S)-Pantolactone can be resolved directly using chiral phase chromatography (28), or after derivatization using optically active acids or isocyanates (29). The sodium salt of pantoic acid is enantioselectively protonated in the presence of (1S)-(+)-10-camphorsulfonic acid and then undergoes spontaneous cyclization to (R)-pantolactone (9) (30).

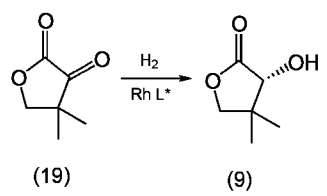
Ketopantolactone (19) is conveniently prepared by oxidation of (*R,S*)-pantolactone (18). Various oxidizing agents have been patented for the oxidation of pantolactone, such as MnO₂ (31), DMSO–Ac₂O (32), and hypohalites (33). An improved yield of ketopantolactone (19) via electrolytic oxidation of pantolactone with an aqueous solution containing an alkali metal salt was reported (34). Ketopantolactone (19) has been prepared in good yield via cyclocondensation of the 2-keto-3-methylbutyrate (20) with formaldehyde (35).



Asymmetric hydrogenation of ketopantolactone (19) in the presence of chiral dirhodium complexes gave (*R*)-pantolactone (9) in high yield and excellent selectivity (36) (Table 2).

Table 2. Asymmetric Hydrogenation of Ketopantolactone (19)

Catalyst	Chiral ligand	Isolated yield %	ee %	Reference
[Rh(bppm)(1,5-hexadiene)Cl]		93	56	37
[Rh(bppm)(COD)Cl]		93	86.7	38
dimethyl-(<i>R</i>)-malate		40	>99	39
di-(μ-carboxylato)bis(aminophosphine-phosphinite) dirhodium complexes		100	91	36



The microbial reduction of ketopantoate (21) to (*R*)-pantoate (22) has been investigated, employing one or two microorganisms (40–44). Reduction with *Proteus vulgaris* gave (*R*)-pantoate (22) in high (460 mmol) concentration, excellent selectivity (97% ee), and 98.5% yield with acceptable productivity numbers. Productivity number (PN) is derived by dividing the concentration in mmol of product formed by the dry weight in kg of the cells, multiplied by the time in hours required for the conversion. Higher PN is favored for production (44) (Table 3).

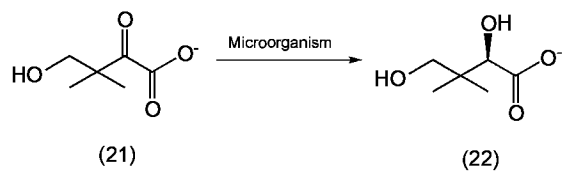


Table 3. Microbial Reduction of Ketopantoate (21) to (*R*)-Pantoate (22)

Microorganism	% Conversion yield	% ee	Final product concentration, mM	PN ^a	Reference
<i>Rhodotorula erythropolis</i>	90.5	94.4	140		40
<i>Agrobacterium</i> sp. S-242	90	>98	910	120	41
<i>Nocardia asteroides</i> / <i>Agrobacterium radiobacter</i>	89	82.8	620	130	42
<i>Candida macedonienis</i> -AK4 4588	97.2	98	80		43
<i>Proteus vulgaris</i>	98.5	>97	460	740	44

^a Productivity number (PN) = product in mmol / dry wt of cells in kg × time, h.

Reduction of ketopantolactone (19) to (*R*)-pantolactone (9) also was evaluated using microbes (45–52) (Table 4).

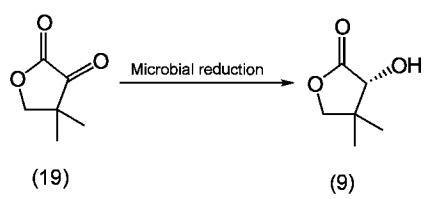


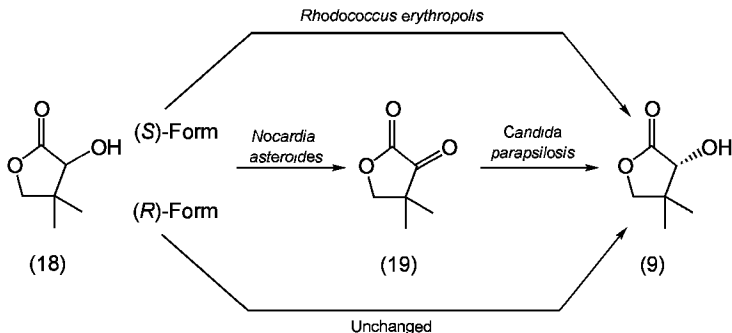
Table 4. Microbial Reduction of Ketopantolactone (19) to (*R*)-Pantolactone (9)

Microorganism	% Conversion yield	% ee	Final product concentration, mM	PN	Reference
<i>Saccharomyces cerevisiae</i>	>90	>98	40	30	45
<i>Byssoschamys fulva</i>	>90	>96			46
<i>Nocardiaasteroides/ candida parapsilosis</i>	>90	>98	390	130	47
<i>Candida parapsilosis</i>	>90	94–98	700		48

<i>Rhodotorulaminuta</i>	80.3	99.7	380	49
<i>Saccharomyces cerevisiae</i>	39	93	20	50
<i>Zygosaccharo-myces</i>	96.8	90		51
Baker's yeast β-cyclodextrin	39	93		52

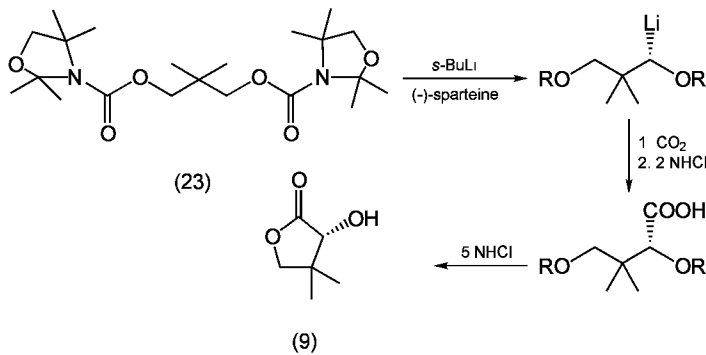
These microbial methods have afforded excellent ee values and yields (40–52). However, owing to low, ie, 30–130, productivity numbers these processes have not yet been commercialized.

In a novel approach, enantiomerically enriched (R)-pantolactone (9) is obtained in a enzymatic two-step process starting from racemic pantolactone.

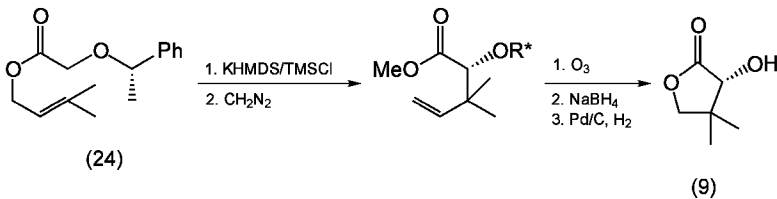


In a first step, *Nocardia asteroides* selectively oxidizes only (S)-pantolactone to ketopantolactone (19), whereas the (R)-pantolactone remains unaffected (47). The accumulated ketopantolactone is stereospecifically reduced to (R)-pantolactone in a second step with *Candida parapsilosis* (product concentration 72 g L⁻¹, 90% molar yield and 100% ee) (48). Racemic pantolactone can also be converted to (R)-pantolactone by one single microbe, ie, *Rhodococcus erythropolis*, by enantioselective oxidation to (S)-pantolactone and subsequent stereospecific reduction in 90% yield and 94% ee (product concentration 18 g L⁻¹) (40).

Although not of industrial importance, several asymmetric syntheses of (R)-pantolactone (9) have been developed. Stereoselective abstraction of the *si*-proton of the achiral 1,3-propanediol derivative (23) by *sec*-butyllithium–(–)-sparteine, followed by carboxylation and hydrolysis, results in (R)-pantolactone in 80% yield and 95% ee (53).

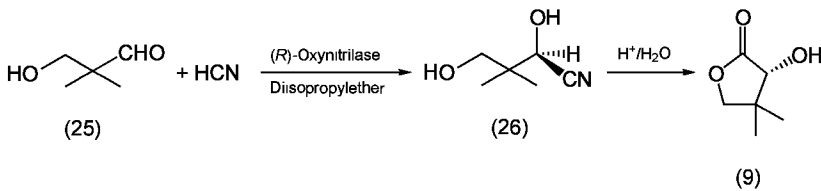


(R)-Pantolactone is also prepared in a sequence involving Claisen rearrangement of the chiral glycolate (24), although with poor enantioselectivity (54).



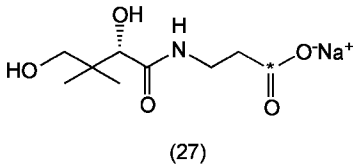
By employing Sharpless epoxidation as a key step, a multistep chemical synthesis of (R)-pantolactone has also been reported (55).

Enantioselective addition of hydrogen cyanide to hydroxypivaldehyde (25), catalyzed by (R)-oxynitrilase, afforded (R)-cyanohydrin (26) in good optical yield. Acid-catalyzed hydrolysis followed by cyclization resulted in (R)-pantolactone in 98% ee and 95% yield after one recrystallization (56).



Labeled Compounds

Various labeled degradation products of pantothenic acid and coenzyme A are known. Both (1-¹⁴C)-sodium pantothenate (27) and (R)-(1-¹⁴C)-panthenol are

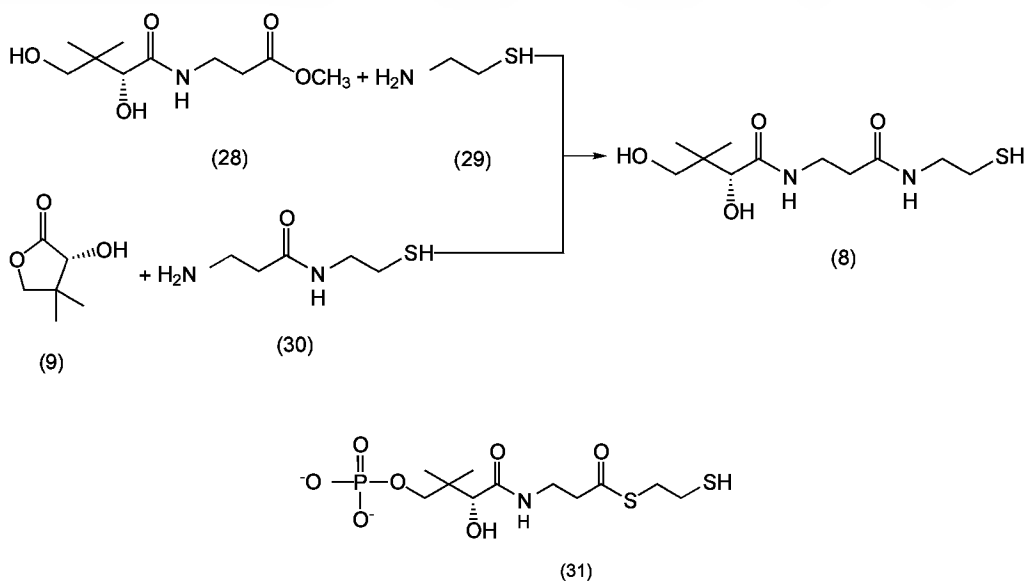


known and are commercially available from New England Nuclear Corporation (NEN, Boston, Mass.). The syntheses of ¹⁴C-phosphopantotheine (57)

and ¹⁴C-labeled coenzyme A (58) have also been reported.

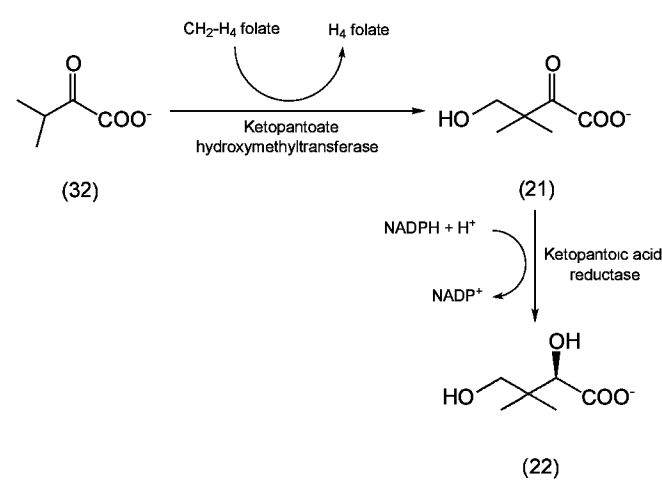
Derivatives and Analogues

Methyl and diacetyl derivatives of pantothenic acid are prepared via methylation and acetylation of pantothenic acid (59). Pantethein (8) is one of the biochemical degradation products derived from coenzyme A. Synthetically, it can be prepared in several different ways, starting from either methyl pantothenate (28) or (R)-pantolactone. Pantethein (8) is prepared either by condensing cysteamine (29) with methyl pantothenate (28) (60) or by coupling (R)-pantolactone (9) with 3-amino-*N*-(2-mercaptoethyl)propanamide (30) (61). Phosphopantetheine analogue (31) was prepared starting from (R)-pantothenic acid. Enzymatic condensation of this compound with adenosin-3,5-diphosphate (6) gave CoA analogue (2b) (62).

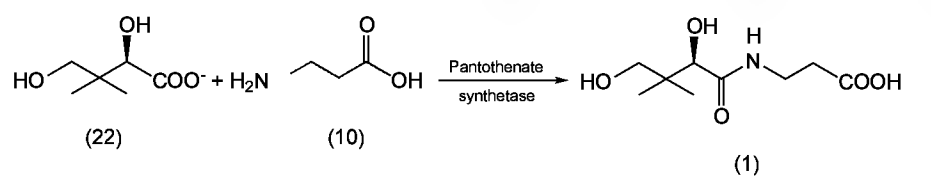


Biosynthesis

The metabolically active form of pantothenic acid is coenzyme A. Pantothenic acid is produced only by microorganisms, starting from (R)-pantoate (22) and β-alanine. (R)-Pantoate is synthesized by a set of enzymatic reactions, as follows (63,64):



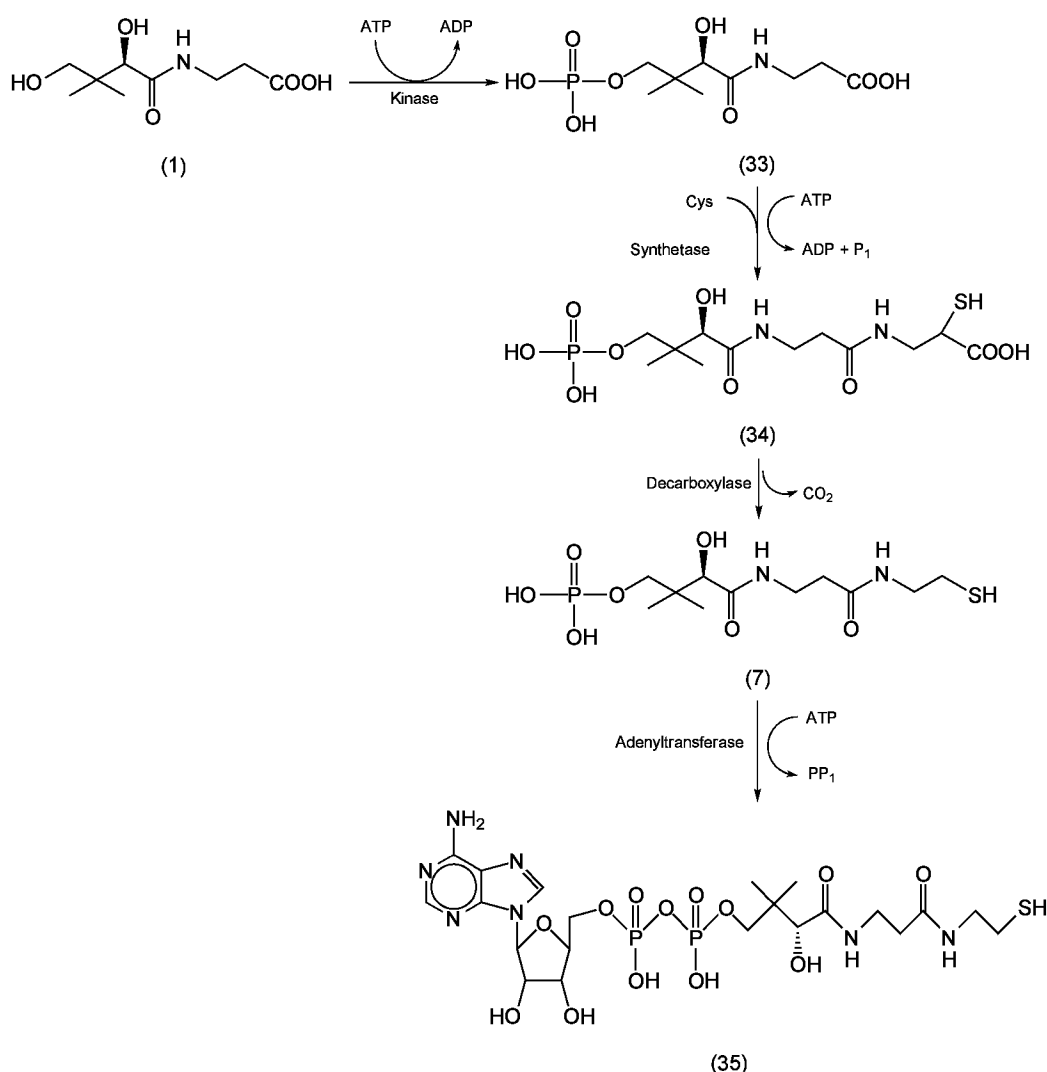
The conversion of α-ketoisovalerate (32) to ketopantoate (21) is catalyzed by ketopantotate hydroxymethyltransferase and a cofactor tetrahydrofolate (65). Further reduction of ketopantoate (21) to (R)-pantoate (22) is catalyzed by ketopantoic acid reductase (66). Aspartic acid decarboxylase catalyzes the decarboxylation of aspartic acid to yield β-alanine (10), a precursor for the biosynthesis of pantothenic acid (67). Finally, (R)-pantothenic acid is obtained by coupling β-alanine (10) with (R)-pantoate (22) in the presence of pantothenate synthetase:



Various pathways have been proposed for the conversion of pantothenic acid to coenzyme A (68,69). The currently accepted pathway involves the following sequence; pantothenate to 4-phosphopantothenic acid (33), 4-phosphopantothenyl cysteine (34), 4-phosphopantetheine (7), dephosphocoenzyme A (35), and coenzyme A (2). Phosphorylation of pantothenic acid to 4-phosphopantothenic acid is catalyzed by pantothenic kinase in the rate-limiting step in the overall synthesis of coenzyme A (70).

Analytical Methods

Chemical, physical, animal, microbiological, and biochemical assays have been used to determine pantothenic acid. Pantothenic acid in foods is found in both the free form and as the vitamin moiety of coenzyme A and phosphopantetheine.



For most assays, the incorporated pantothenic acid has to be liberated enzymatically. Usually, a combination of pantotheinase and alkaline phosphatase is used to liberate the bound pantothenic acid. The official method for pantothenic acid of the Association of Official Analytical Chemists (AOAC) is the microbiological assay that uses *L. Plantarium* (ATCC 8014) as the test organism (71). Samples are extracted at 121°C at pH 5.6–5.7, proteins are precipitated at pH 4.5, and the resulting clear extracts are adjusted to pH 6.8 prior to assay. This procedure is only suitable to determine calcium pantothenate or other free forms of pantothenic acid.

A radioimmunoassay method has been developed by Wyse and co-workers (72). It is a sensitive method for the determination of small amounts of pantothenic acid in biological fluids. The assay is based on the binding of an enzyme specific for pantothenic acid. Hemolyzed blood samples are subjected to a double-enzyme treatment with 5 IU of bovine intestinal alkaline phosphatase and 0.1 IU of pantotheinase. After hydrolysis, the clear supernatant extract is analyzed for pantothenate by a radioimmunoassay method. In the case of pharmaceutical preparations, vitamin concentrations are determined by using spectrophotometric and fluorometric methods (73). For spectrophotometric methods, detection is done at 358 nm and for fluorescence, excitation is done at 350 nm and collection at 450 nm. Food pantothenate assays involve hydrolysis of the food by 25% hydrochloric acid in a 95–100°C water bath, extraction of the pantoic lactone into dichloromethane and analysis by gas chromatography with flame ionization detection (74). Chiral polysiloxane fused silicon capillary columns (XE-60-L-Val-(*S*)- or (*R*)- α -phenylethylamide) have been developed for separation of the (*R*)- and (*S*)-pantothenic acid mixture (75). Recently, clinical separation and simultaneous determination of (*R*)- and (*S*)-pantothenic acids in rat plasma using gc–ms has been described (76). Deproteinization of the plasma samples is done by eluting the plasma sample through an anion-exchange resin using 1 *M* sodium chloride solution. Pantothenic acid is extracted quantitatively with ethyl acetate after acidifying the basic aqueous solution. After esterification of the carboxylic group, each derivative is identified using gc–ms.

Deficiency

A deficiency of pantothenic acid in humans has not been detected because it is widely available in food (77). In the case of prisoners of war in the 1940s, a burning feet syndrome, which was attributed to pantothenic acid deficiency as a result of severe malnutrition, was observed (78). Pantothenic acid deficiency in humans is induced experimentally by feeding a diet deficient in pantothenic acid along with a pantothenic acid antagonist such as ω -methylpantothenic acid. This results in clinical malaise, fatigue, headache, sleep disturbance, nausea, vomiting, and cardiovascular instability. Impaired responses to insulin, histamine, and ACTH (stress hormone) have also been observed. A wide variety of abnormalities such as retarded growth, impaired fertility, gastrointestinal lesions, and adrenal necroses are reported when rats are fed a pantothenic acid-deficient diet (79,80). Chicks on a pantothenic acid-deficient diet develop lesions at the corners of the mouth, swollen eyelids, hemorrhagic cracks on the feet, and listlessness (81).

Absorption

Pantothenic acid occurs in most foods and feedstuffs as CoA and acyl-carrier protein. The utilization of the vitamin depends upon the hydrolytic digestion of these protein complexes to release the free vitamin. Coenzyme A is hydrolyzed in the intestinal lumen before absorption into the cell by a sodium ion-dependent, passive mechanism (82,83). High concentrations of pantothenic acid are found in the heart and kidney, compared to other organs such as the intestine and liver (84). Pantothenic acid utilization in the formation of acetyl CoA is impaired in alcoholics because the ethanol metabolite acetaldehyde inhibits the conversion (85). Erythrocytes, which carry most of the vitamin in the blood, carry it predominantly in the form of CoA. Blood and urinary levels of pantothenic acid are lower for oral contraceptive users than for nonusers (86,87). Absorption is also inhibited in the presence of known antagonists, such as ω -methylpantothenic acid, and to some extent with (*S*)-pantothenic acid. Another factor which influences vitamin absorption is a high fat diet. This condition significantly lowers the utilization of pantothenic acid in the formation of coenzyme A in the liver because of changes in lipid metabolism (88). On the other hand, high levels of protein in the diet may promote better utilization of this vitamin in the synthesis of coenzyme A (89).

Nutritional Requirements

Pantothenic acid is widely distributed in food and because of the lack of conclusive evidence regarding quantitative needs, a recommended dietary allowance (RDA) for pantothenic acid has not been established. In 1989, the Food and Nutrition Board of the United States National Research Council suggested a safe intake of 4–7 mg d for adults. The provisional allowance for infants is 2–3 mg daily (90).

Functions

The most important functions of pantothenic acid are its incorporation in coenzyme A and acyl carrier protein (ACP). Both CoA and ACP 4-phosphopantetheine function metabolically as carriers of acyl groups. Coenzyme A forms high-energy thioester bonds with carboxylic acids. The most important coenzyme is acetyl CoA. Acetic acid is produced during the metabolism of fatty acids, amino acids, or carbohydrates. The active acetate group of acetyl CoA can enter the Krebs cycle and is used in the synthesis of fatty acids or cholesterol. ACP is a component of the fatty acid synthase multienzyme complex. This complex catalyzes several reactions of fatty acid synthesis (condensation and reduction). The nature of the fatty acid synthase complex varies considerably among different species (91).

Toxicity

Pantothenic acid toxicity has not been reported in humans. Massive doses (10 g d) in humans have produced mild intestinal distress and diarrhea. Acute toxicity was observed in case of mice and rats by using calcium pantothenate at fairly large doses (92).

Uses

The bulk of the industrial supply of the calcium salt of (R)-pantothenic acid is used in food and feed enrichment. Food enrichment includes breakfast cereals, beverages, dietetic, and baby foods. Animal feed is fortified with calcium-(R)-pantothenate which functions as a growth factor. Panthenol is used in skin care products. When applied topically to alleviate itching, it helps to keep the skin moist and supple, stimulates cell growth, and accelerates wound healing by increasing the fibroblast content of scar tissue (93,94). It is also used as a moisturizer and conditioner in hair care products. It is believed to protect hair against chemical and mechanical damage caused by perming, coloring, and shampooing (95).

Economic Aspects

The major producers of the calcium salt of pantothenic acid and panthenol are Hoffmann-La Roche, Daiichi, BASF, and Alps. Racemic panthenol is used mainly in hair care products, whereas (R)-panthenol is exclusively used in top-of-the-line, more expensive skin and hair care products. The current (ca 1997) price of (R,S)-panthenol varies between \$12 and \$18 per kg, that of D-panthenol varies between \$30 and \$45 per kg, and that of D-calcium pantothenate (Calpan) varies between \$22 and \$30 per kg.

Future Prospects

Despite the progress made in the stereoselective synthesis of (R)-pantothenic acid since the mid-1980s, the commercial chemical synthesis still involves resolution of racemic pantolactone. Recent (ca 1997) synthetic efforts have been directed toward developing a method for enantioselective synthesis of (R)-pantolactone by either chemical or microbial reduction of ketopantolactone. Microbial reduction of ketopantolactone is a promising area for future research.

BIBLIOGRAPHY

"Pantothenic Acid" in *ECT* 2nd ed., Vol. 9, pp. 805–811, by R. J. Williams, University of Texas, and W. Wenner, Hoffmann-La Roche Inc.

1. R. J. Williams, C. M. Lyman, G. H. Goodyear, J. H. Truesdail, and H. Holaday, *J. Am. Chem. Soc.* **55**, 2912 (1933).
2. D. W. Woolley, H. A. Waisman, and C. A. Elvehjem, *J. Biol. Chem.* **129**, 673 (1939).
3. T. H. Jukes, *J. Am. Chem. Soc.* **61**, 975 (1939).
4. E. T. Stiller, S. A. Harris, J. Finkelstein, J. C. Keresztesy, and K. Folkers, *J. Am. Chem. Soc.* **62**, 1785 (1940).
5. T. Reichstein and A. Grössner, *Helv. Chem. Acta*, **23**, 650 (1940).
6. R. Kuhn and Th. Wieland, *Chem. Ber.* **73**, 962 (1940).
7. J. Baddiley and E. M. Thain, *J. Chem. Soc.*, 3421 (1951); J. Baddiley and E. M. Thain, *J. Chem. Soc.*, 1610 (1953).
8. M. L. Orr, *Home Economic Research Report* No. 36, U.S. Department of Agriculture, Washington, D.C., 1969.
9. J. H. Walsh, B. W. Wyse, and R. G. Hansen, *J. Am. Diet. Assoc.* **78**, 140 (1981).
10. J. B. Tarr, T. Tamura, and E. L. R. Stokstad, *Am. J. Clin. Nutr.* **34**, 1328 (1981).
11. IUPAC-IUB Joint commission on biochemical nomenclature (JCBN), *Pure Appl. Chem.* **59**, 834 (1987).
12. J. Marks, *The Vitamins in Health and Disease*, Little, Brown & Co., Boston, MA, 1968, pp. 121–125.
13. D. Novelli, N. O. Kaplan, and F. Lipman, *Fed. Proc.* **9**, 209 (1950).
14. O. Isler, G. Brubacher, S. Ghisla, and B. Kautler, *Vitamine II*, Georg Thieme Verlag, New York, 1988, p. 309.
15. E. H. Wilson, J. Weiglard, and M. Tischler, *J. Am. Chem. Soc.* **76**, 5177 (1954); **Ger. Pat. DE 2919515** (Nov. 22, 1979), D. Bartoldus and E. Broger (to Hoffmann-La Roche Inc.).
16. O. Schnider in *Festschrift Emil Borell*, Hoffmann-La Roche, Basel, Switzerland, 1946, p. 85; E. E. Snell and W. Shive, *J. Biol. Chem.* **158**, 551 (1945); G. S. Kozłowa, T. I. Erokchina, and V. I. Gunar, *Pharm. Chem. J.* **11**, 505 (1977).
17. J. Baddiley and A. P. Mathias, *J. Chem. Soc.*, 2803 (1954).
18. **Ger. Pat. DE 2758883** (July 5, 1979), H. Distler and W. Goetze (to BASF); E. Glaser, *Herstellung Monatsch. Chem.* **25**, 46 (1904); **Jpn. Pat. 55062080** (May 10, 1980) (to UBE Industries Ltd.).
19. S. Shimizu and H. Yamada, in T. O. Baldwin, F. M. Raushel, and A. I. Scott, eds., *Chemical Aspects of Enzyme Biotechnology*, Plenum Press, New York, 1990, p. 151.
20. **Ger. Pat. DE 2404305** (Oct. 27, 1983), S. Pfohl and co-workers (to BASF AG).
21. **Jpn. Pat. 02240073**, (Sept. 25, 1990) S. Takeda, M. Takeda, K. Suzuki, and M. Yuya (to Fuji Chemical Industrial Co., Ltd.).
22. **Ger. Pat. DE 2558508**, (Sept. 7, 1976) A. Kinugasa, T. Okuda, M. Goto, and M. Saito (to Alps Pharmaceutical Industrial Co., Ltd.).
23. **Jpn. Pat. 8115240** (Aug. 10, 1982) (to Sumitomo Chemical Co., Ltd.).
24. C. Fizet, *Helv. Chim. Acta*. **69**, 404 (1986).
25. **Jpn. Pat. 8735340**, (Aug. 23, 1988) H. Nohira, S. Yoshida, M. Nohira, K. Sato (to Daicel Chemical Industries Ltd.).
26. F. Toda, A. Sato, K. Tanaka, and T. C. W. Mak, *Chem. Lett.*, 873 (1989).
27. I. Masahiro, *Res. Inst. Yakugaku Zasshi* **96**, 71 (1976); **Jpn. Pat. 47022220**, (June 22, 1972) S. Nabeta, Y. Nakabe, and M. Nagaki (to Daiichi Seiyaku Co., Ltd.).
28. N. Oi, T. Doi, H. Kitahara, and Y. Inda, *J. Chromatogr.* **208**, 404 (1981); W. A. Kunig and U. Sturm, *J. Chromatogr.* **328**, 357 (1985); E. Francotte and D. Lohmann, *Helv. Chim. Acta*. **70**, 1569 (1987); D. W. Armstrong, W. Li, C. Chang, and J. Pitha, *Anal. Chem.* **62**, 914 (1990); W.-Y. Li, H. L. Jin, and D. W. Armstrong, *J. Chromatogr.* **509**, 303 (1990).
29. T. Arai, H. Matsuda, and H. Oizumi, *J. Chromatogr.* **474**, 405 (1989); A. Takasu and K. Ohya, *J. Chromatogr.* **389**, 251 (1987).
30. K. Fuji, M. Node, M. Murata, S. Terada, and K. Hashimoto, *Tetrahedron Lett.* **27**, 5381 (1986).

31. **Jpn. Pat. 04095087** A2, (Mar. 27, 1992) N. Kuroda and K. Kashiwa (to Takeda Yakuhin Kogyo K. K.).

32. **Jpn. Pat. 04095086** A2, (Mar. 27, 1992) N. Kuroda and K. Kashiwa (to Takeda Yakuhin Kogyo K. K.).

33. **Jpn. Pat. 04095084** A2, (Mar. 27, 1992) N. Kuroda and K. Kashiwa (to Takeda Yakuhin Kogyo K. K.).

34. **Jpn. Pat. 01208488** A2, (Aug. 22, 1989) H. Sato, S. Takeda, and M. Yuya (to Fuji Chemical Industrial Co., Ltd.).

35. DE 3229026 A1, (Aug. 4, 1982) K. Halbritter and M. Eggersdorfer (to BASF AG) (Aug. 4, 1982).

36. J.-F. Carpentier, F. Agbossou, and A. Mortreux, *Tetrahedron Asymm.* **6**, 39 (1995).

37. K. Achiwa, T. Kogure, and I. Ojima, *Tetrahedron. Lett.* **18**, 4431 (1977).

38. T. Ojima, T. Kogure, and Y. Yoda, *Org. Synth.* **63**, 18 (1985).

39. I. Ojima, T. Kogure, and T. Terasaki, *J. Org. Chem.* **43**, 3444 (1978).

40. S. Shimizu, S. Hattori, H. Hata, and H. Yamada, *Appl. Environ. Microbiol.* **53**, 519 (1987).

41. M. Kataoka, S. Shimizu, and H. Yamada, *Agric. Biol. Chem.* **54**, 177 (1990).

42. M. Kataoka, S. Shimizu, and H. Yamada, *Recl. Trav. Chim. Pays-Bas.* **110**, 155 (1991).

43. M. Kataoka, S. Shimizu, Y. Doi, and H. Yamada, *Appl. Environ. Microbiol.* **56**, 3595 (1990).

44. R. Eck and H. Simon, *Tetrahedron Asymm.* **5**, 1419 (1994).

45. R. Kuhn and Th. Wieland, *Chem. Ber.* **75**, 121 (1942).

46. R. P. Lanzilotta, D. G. Bradley, and K. M. McDonald, *Appl. Microbiol.* **27**, 130 (1974).

47. S. Shimizu, S. Hattori, H. Hata, and H. Yamada, *Enzyme Microb. Technol.* **9**, 411 (1987).

48. H. Yamada and S. Shimizu in D. S. Clark, D. Estell, and J. Dordick, eds., *Enzyme Engineering*, Vol. 11, The New York Academy of Sciences, New York, 1992, p. 374.

49. S. Shimizu, H. Yamada, H. Hata, T. Morishita, S. Akutsu, and M. Kawamura, *Agric. Biol. Chem.* **51**, 289 (1987).

50. K. Nakamura, S.-I. Kondo, Y. Kawai, and A. Ohno, *Tetrahedron Asymm.* **4**, 1253 (1993).

51. **Jpn. Pat. 05227987** A2, (Sept. 7, 1993) K. Sakamoto, H. Yamada, and A. Shimizu (to Fuji Yakuhin Kogyo KK).

52. **Jpn. Pat. 06311889** A2, (Apr. 30, 1993) J. Ohno, K. Nakamura, K. Yasushi and K. Shinichi (to Kokai Tokkyo Koho); K. Nakamura, K. Shin-ichi, and O. Atsuyoshi, *Bioorganic. Med. Chem.* **2**, 433 (1994).

53. M. Paetow, H. Ahrens, and D. Hoppe, *Tetrahedron Lett.* **33**, 5323 (1992).

54. J. Kallmerten and T. Gould, *J. Org. Chem.* **51**, 1152 (1986).

55. A. V. Rama Rao, S. Mahender Rao, and G. V. M. Sharma, *Tetrahedron Lett.* **35**, 5735 (1994).

56. F. Effenberger, J. Eichhorn, and J. Roos, *Tetrahedron Asymm.* **6**, 271 (1995).

57. C. J. Chesterton, P. H. W. Butterworth, and J. W. Porter, *Meth. Enzymol.* **18A**, 371 (1970).

58. C. J. Chesterton, P. H. W. Butterworth, and J. W. Porter, *Meth. Enzymol.* **18A**, 364 (1970).

59. **U.S. Pat. 2441949** (May 25, 1948) S. H. Babcock (to the University of California, Berkeley, Calif.).

60. E. E. Snell and co-workers, *J. Am. Chem. Soc.* **72**, 5349 (1950).

61. J. Baddiley and E. M. Thain, *J. Chem. Soc.*, 800 (1952).

62. D. P. Martin and D. G. Drueckhammer, *J. Am. Chem. Soc.* **114**, 7288 (1992).

63. D. S. Vallari and C. O. Rock, *J. Bacteriol.* **162**, 1156 (1985).

64. G. M. Brown and J. M. Williamson, *Adv. Enzymol.* **53**, 345 (1982).

65. J. H. Teller, S. G. Powers, and E. E. Snell, *J. Biol. Chem.* **251**, 3780 (1976).

66. S. Shimuzu, M. K. Qtaoka, M. C. M. Chunag, and H. Yamandu, *J. Biol. Chem.* **263**, 1207 (1988).

67. J. M. Williamson and G. M. Brown, *J. Biol. Chem.* **254**, 8074 (1979).

68. D. G. Novelli, *Physiol. Rev.* **33**, 523 (1953).

69. Y. Shigeta and M. Shichiri, *J. Vitamin.* **12**, 186 (1966).

70. Y. Abiko, S. Ashida, and M. Shimuzu, *Biochim. Biophys. Acta* **268**, 364 (1972).

71. D. Bhatia, ed., in *Encyclopedia of Food Science and Technology*, John Wiley & Sons, Inc., New York, 1991, p. 2783; B. W. Wyse, in J. Augustin, B. P. Klein, D. A. Becker, and P. B. Venugopal, eds., *Methods of Vitamin Assay*, 4th ed., John Wiley & Sons, Inc., New York, 1985, Chapt. 16, p. 2783.

72. B. W. Wyse, C. Withwer, and R. G. Hansen, *Clin. Chem.* **25**, 108 (1979).

73. R. B. Roy and A. Buccafuri, *J. Assoc. Off. Anal. Chem.* **61**, 720 (1978).

74. J. Davidek, J. Velisek, J. Cerna, and T. Davidek, *J. of Micronutr. Anal.* **1**, 39 (1985).

75. W. A. Kunig and U. Sturm, *J. Chromatogr.* **328**, 357 (1985).

76. K. Banno, S. Horimoto, and M. Matsuoka, *J. Chromatogr.* **564**, 1 (1991).

77. R. E. Hodges, M. A. Ohlson, and W. B. Bean, *J. Clin. Invest.* **37**, 1642 (1958).

78. M. Glusman, *Am. J. Med.* **3**, 211 (1947).

79. E. Kazuko, N. Kubato, T. Nishigaki, and M. Kikutani, *Chem. Pharm. Bull.* **23**, 1 (1975).

80. M. M. Nelson, F. Van Nouheys, and H. M. Evans, *J. Nutr.* **34**, 189 (1947).

81. D. M. Hegsted, J. J. Oelson, R. C. Mills, C. A. Elvehjem, and E. B. Hart, *J. Nutr.* **20**, 599 (1940).

82. G. D. Lopaschuk, M. Michalak, and H. Tsang, *J. Biol. Chem.* **262**, 3615 (1987).

83. K. Shibata, C. J. Gross, and L. M. Henderson, *J. Nutr.* **113**, 2107 (1983).

84. D. K. Reibel, B. W. Wyse, D. A. Berkich, and J. R. Neely, *Am. J. Physiol.* **240**, H 606 (1981).

85. B. R. Eissenstat, B. W. Wyse, and R. G. Hansen, *Am. J. Clin. Nutr.* **44**, 931 (1986).

86. W. B. Bean and R. E. Hodges, *Proc. Soc. Exp. Biol. Med.* **86**, 693 (1954).

87. P. C. Fry, H. M. Fox, and H. G. Tao, *J. Nutr. Sci. Vitaminol.* **22**, 339 (1976).

88. Y. Furukawa and S. Kimura, *J. Vitamin.* **18**, 213 (1972).

89. R. W. Luecke, J. A. Hoefler, and F. Thorp, Jr., *J. Anim. Sci.* **11**, 138 (1952).

90. Committee on Dietary Allowances, Food and Nutrition Board, *Recommended Dietary Allowances*, 10th ed., National Academy Press, Washington, D.C., 1989.

91. G. F. Combs, *The Vitamins*, Academic Press, San Diego, Calif., 1992, Chapt. 15, p. 345.

92. K. Unna and J. C. Greslin, *J. Pharmacol. Exp. Ther.* **73**, 85 (1941).

93. J. F. Grenier, M. Aprahamian, C. Genot, and A. Dentinger, *Acta Vita. et Enzy.* **4**, 81 (1982).

94. M. Aprahamian, A. Dentinger, C. Stoch-Damge, J. C. Kouassi, and J. F. Grenier, *Am. J. Clin. Nutr.* **41**, 578 (1985).

95. F. Vaxman and co-workers, *Eur. Surg. Res.* **27**, 158 (1995).

General References

D. Bhatia, ed., Vitamin, Part XIV: Pantothenic acid in *Encyclopedia of Food Science and Technology*, John Wiley & Sons, Inc., New York, 1991, p. 2783.

H. M. Fox in L. J. Machlin, eds., *Handbook of Vitamins*, Marcel Dekker Inc., New York, 1991, pp. 429–451.

O. Isler, G. Brubacher, S. Ghisla, and B. Kautler, *Vitamine II*, Georg Thieme Verlag, New York, 1982, pp. 309–339.

S. Shimizu and Y. Yamada, in T. O. Baldwin, F. M. Raushel, and A. I. Scott, eds., *Chemical Aspects of Enzyme Biotechnology*, Plenum Press, New York, 1990, p. 151.

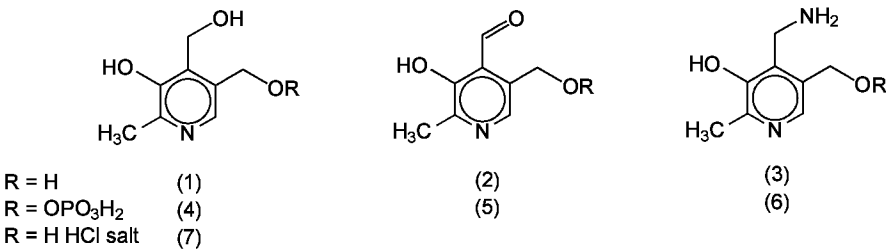
B. W. Wyse, in J. Augustin, B. P. Klein, D. A. Becker, and P. B. Venugopal, eds., *Methods of Vitamin Assay*, 4th ed., John Wiley & Sons, Inc., New York,

1985, Chapt. 16.
G. F. Combs, *The Vitamins*, Academic Press, San Diego, Calif., 1992, Chapt. 15, pp. 345 ff.

Thimma R. Rawalpally
Hoffmann-La Roche Inc.

PYRIDOXINE (B₆)

Pyridoxine [65-23-6] (vitamin B₆) is the recommended IUPAC-IUB designation for a group of closely related, biologically interconvertible 3-hydroxy-2-methylpyridines that exhibit the biological activities of pyridoxine in rats (1). This group includes pyridoxine [65-23-6] (1), pyridoxal [66-72-8] (2), and pyridoxamine [85-87-0] (3) and their respective 5'-phosphates (4–6). In common usage, though, the term vitamin B₆ most often refers to pyridoxine hydrochloride [58-56-0] (7), the most important commercial form (Fig. 1). Contrary to previous proposals and explanations, pyridoxine is not to be used as a general name for this group of vitamins. Pyridoxine is the official name of the specific substance 5-hydroxy-6-methyl-3,4-bis(hydroxymethyl)pyridine (1) which usually comprises only a fraction of the total amount of vitamin B₆ in natural sources. The older term pyridoxol is seldom used.



pyridoxal	(2)	[66-72-8]	C ₈ H ₉ NO ₃	167.16			50	4.2, 8.7, 13
pyridoxal HCl		[65-22-5]	C ₈ H ₉ NO ₃	203.63	rhombic	173 dec	50	
pyridoxal-5'-PO ₄	(5)	[54-47-7]	C ₈ H ₁₀ NO P	247.14			soluble	3.5, 8.4

The B vitamins are labile. Pyridoxine is the most stable and pyridoxal the least. The reactions of the substituents at the 4-carbon dominate. On mild oxidation, pyridoxine is converted to pyridoxal (2), which is in equilibrium with its hemiketal [17281-92-4] (8) (13). Air oxidation gives pyridoxic acid [82-82-6] (9), an intensely blue fluorescent substance of value for quantitation (4). Heating pyridoxine with ammonia converts it to pyridoxamine (3) and with alcohols to the 4-ethers (13). Pyridoxine, pyridoxal, and pyridoxamine are relatively stable in acid when cold or heated at 100°C for short periods. Prolonged heating of pyridoxine in hydrochloric acid gives the 4,5-bis-chloromethyl analogue, the 4-chloromethyl group of which is preferentially hydrolyzed. Heating pyridoxine in neutral solution in the absence of good nucleophiles gives dimeric products (14). Some of these reactions are readily rationalized by an intermediate quinone-methide (10) (Fig. 2).

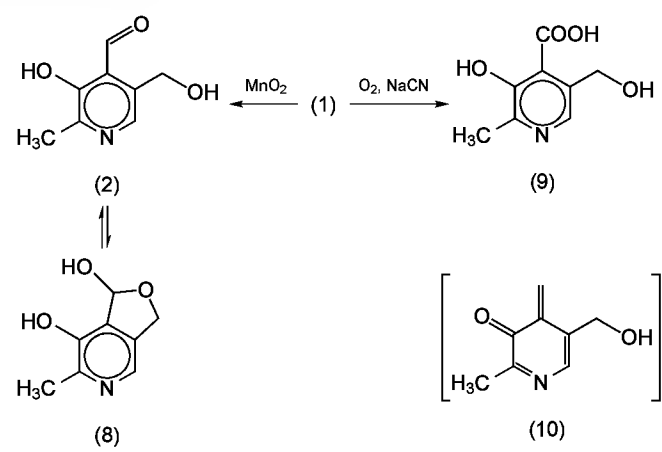


Fig. 2. Reactions of pyridoxal (2).

Pyridoxal and its 5'-phosphate undergo typical aldehyde reactions, such as aldol condensations, bisulfite addition, etc (4,11,12). Both form imines, assisted by the electronegative 2-methylpyridine ring and the acidic 3-hydroxyl. Extensive kinetic and thermodynamic data on imine formation in model systems are available (4). The epsilon-amino group of a lysine unit is usually involved in the active sites of enzymes but other groups can bind pyridoxal during thermal processing and storage, reducing bioavailability (4,15). Hydroxylamines and hydrazine derivatives, such as isoniazid, react and cause enzyme inhibition (4). Imine formation with pyridoxal is used as a tool for selective modification of proteins in biochemical studies (4).

Reactions of imines and their tautomers are the basis of the important biochemistry of pyridoxal 5'-phosphate (5), the coenzyme form for over 100 enzymatic reactions. Reaction with an epsilon-amino group of a lysyl residue and ionic interaction of the 5'-phosphate binds pyridoxal to the enzyme in the resting state. Displacement of lysine by a substrate amino group initiates a sequence of facile equilibrations of the aldimine, ketimine, and enamine forms, along with elimination and addition reactions which rationalize the various conversions and products. Known reactions involving the alpha-, beta-, and gamma-carbons of the amino acid and their substituents include decarboxylations, transaminations, racemizations, aldol and retroaldol reactions, eliminations, and hydration reactions (4) (Fig. 3). Many of these transformations have been modeled *in vitro* in chemical studies. Some of this chemistry can cause covalent bonding of a substrate to other reactive residues in the active site, resulting in irreversible "suicide inactivation" of the enzyme. This is a mechanism of action of some naturally occurring toxins and a protocol in rational drug design (4).

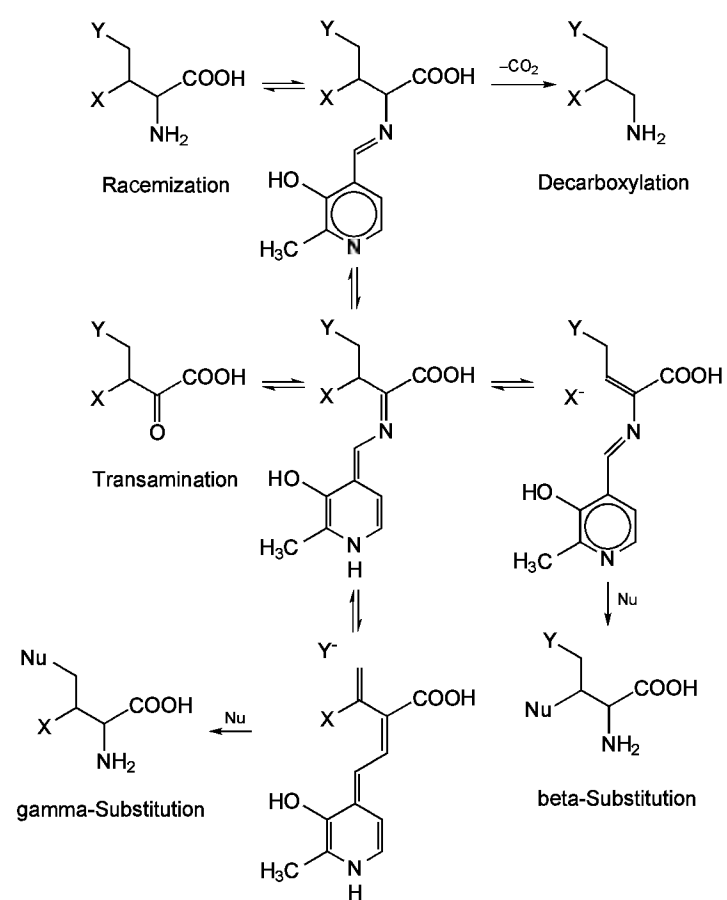


Fig. 3. Reactions of imines and their tautomers.

Pyridoxine hydrochloride is stable if kept dry or in acidic solution. Its aqueous solutions may be heated for 30 min at 120°C without decomposition (16). Losses have been reported for mixtures of the vitamins stored in solution at room temperature or below unless they are kept acidic (17). Decomposition in foods on processing and storage depends mainly on pH, moisture, and temperature. First-order kinetics are often followed and stabilities can be estimated using Arrhenius calculations (4,18). All B vitamins are susceptible to decomposition by near-uv light to give reddish brown materials, especially in solution at neutral or higher pH. Pyridoxal and pyridoxamine decompose most rapidly, even in acid (18,19). Gamma irradiation of food can also cause losses (20). Hydroxyl radical formed from ascorbate gives the 6-hydroxyl derivative of pyridoxine. This process may occur during the storage of fruits and vegetables containing high ascorbate levels (21,22).

Salts and Derivatives. Generally the B vitamins are high melting crystalline solids that are very soluble in water and insoluble in most other solvents. Properties of the common forms are listed in Table 1. The only commercially important form of vitamin B6 is pyridoxine hydrochloride (7). This odorless crystalline solid is composed of colorless platelets melting at 204–206°C (with decomposition). In bulk, it appears white and has a density of ~0.4 kg/L. It is very soluble in water (ca 0.22 kg/L at 20°C), soluble in propylene glycol, slightly soluble in acetone and alcohol (ca 0.014 kg/L), and insoluble in most lipophilic solvents. A 10⁰ % water solution shows a pH of 3.2. Both the hydrochloride and corresponding free base sublime without decomposition (16).

Occurrence

Vitamin B6 occurs in the tissues and fluids of virtually all living organisms, attesting to its essential role in amino acid biochemistry. Pyridoxal phosphate bound to protein is the main form in animals. Pyridoxine, pyridoxamine, and their phosphates are the usual forms in plants (23,24). Significant percentages in plants occur as pyridoxine glycosides, which are substantially less bioavailable (23,25,26). Nearly all foods in a typical human diet contain detectable amounts of the B6 vitamins. Relatively rich sources are edible yeast, meats (especially salmon and calf liver), whole grain cereals, legumes, and nuts (qv). Cheeses with extensive mold growth are also good sources. Broccoli, cauliflower, spinach, corn, potatoes, bananas, and raisins are good sources, but vegetables and fruits in general are not (24) (Table 2).

Table 2. Average Vitamin B6 Contents of Raw Foods, mg/100 g^a

Food	Vitamin B6	Food	Vitamin B6
beef liver	0.84	brown rice	0.55
beef	0.33	white rice	0.17
chicken	0.33–0.68	whole wheat flour	0.34
pork	0.35	white flour	0.06
fish	0.17–0.43	beans	0.56–0.81
eggs	0.11	peanuts	0.40
whole milk	0.04	corn	0.25
bananas	0.51	potatoes	0.25
raisins	0.24	vegetables, green	0.15–0.28
other fruits	0.02–0.07	tomatoes	0.10

^a By microbial assay.

As vitamin B6 is mainly located in the germ and aleurone layer in cereal grains; polishing for the production of flour removes a substantial portion. White bread is therefore a poor source unless fortified. Some nonedible yeasts contain up to 38 mg/100 g dry weight vitamin B6, the highest level of the natural sources (4,27). As a rule, these amounts are too low for cost-effective isolation.

Biochemical and Physiological Functions

Because vitamin B6 is widely distributed in nature, severe deficiency symptoms are seldom observed in humans with adequate diets. However, there is little doubt vitamin B6 is essential in human and animal nutrition. As enzyme cofactors, pyridoxal 5'-phosphate and pyridoxamine 5'-phosphate are required for important biotransformations of amino acids (4). Decarboxylations of amino acids give rise to a number of important amines, among them the vasodilator histamine, the vasoconstrictor and neurotransmitter serotonin, and the major neurotransmitters gamma-aminobutyric acid, epinephrine, norepinephrine, tyramine, and dopamine. Transamination interconverts specific pairs of amino acids and their corresponding ketoacids and amino groups and aldehydes, key functions in amino acid biosynthesis and catabolism. Racemization provides D-amino acids, constituents of the cell walls of bacteria. Other important transformations requiring pyridoxal include bioconversions of tryptophan (involved in the biosynthesis of the vitamin niacin), serine transhydroxymethylation (a key step in one carbon metabolism leading to nucleic acid synthesis), glycogen phosphorylation (in liver and muscle-maintaining glucose supplies), biosynthesis of delta-aminolevulinic acid (a precursor to heme synthesis), and key steps in the metabolism of sulfur-containing amino acids. Other roles of pyridoxal and its phosphate include binding to hemoglobin (affecting oxygen binding affinity), effects on the conversion of linoleic acid to arachidonic acid, possible effects on cholesterol and fatty acid metabolism, and modulation of steroid action. As a result, vitamin B6 levels affect gluconeogenesis, nervous system activity, red cell formation and metabolism, immune function, and hormone functions (4,23).

A large variety of biochemical lesions are found in deficiency (23). Depressed antibody, nucleic acid, and protein synthesis are characteristic. Consequences of vitamin B6 deficiency in humans and many animals include scaling dermatitis (especially around the eyes, nose, and mouth), anemia, loss of appetite, poor growth or weight loss, muscular weakness, central nervous system changes such as irritability or depression, kidney stones, and abnormalities seen by electroencephalography. In severe cases, nerve degradation leading to convulsive seizures and death can occur. Possible vitamin B6 deficiency must be assessed; estimation of dietary intake is not sufficient. An older method measures the level of pyridoxal-dependent metabolite xanthurenic acid in the urine after ingestion of a large dose of tryptophan. Improved methods measure pyridoxal phosphate in plasma or pyridoxic acid excretion in urine by sensitive chromatographic procedures (23).

Humans cannot biosynthesize vitamin B6 and must obtain it from dietary sources. Ruminant animals derive some benefit from microfloral supply in the stomach. All animals interconvert the six major forms (1–6). Thus requirements of the enzyme cofactor forms can be maintained by ingestion of pyridoxine only. The bioavailability of vitamin B6 in foods is still an active research area with some unresolved issues (4,15,23). Reported levels of availability vary widely. Earlier studies suggest processing, storage, and preparation could lead to losses of up to 80% of vitamin B6, depending on conditions. Irreversible reductive binding to lysyl residues is reported to give inactivity or even antivitamin effects (15,28). One study estimates bioavailability of vitamin B6 from foods representing the average North American human diet at ca 70–80% (29).

In humans, vitamin B6 in phosphorylated forms is mostly unavailable until hydrolyzed, then passively absorbed in the small intestine. The free vitamins are interconverted and supplied to the circulation by the liver, then enter cells by simple diffusion and are phosphorylated. A typical adult pool of

pyridoxal is estimated to be about <25 mg, most of it bound to enzymes in muscle (23). Small amounts of all forms are excreted in the urine. Most is oxidized in the liver to the main urinary metabolite 4-pyridoxic acid (9), the corresponding diacid, and ring-cleaved fragments (4,30).

Human requirements for vitamin B₆ parallel protein metabolism, with increased intake of protein requiring higher intake of the vitamin. As little as 0.010–0.015 mg of vitamin B₆ per gram of protein prevents deficiency signs with adequate protein intake. Slightly higher intake maintains acceptable metabolite excretion and plasma vitamin levels. The Recommended Daily Allowance (RDA) for vitamin B₆ assumes consumption of twice the recommended amount of protein (Table 3) (31). Allowances for pregnant and lactating women reflect increased protein needs. Low levels in breast milk place infants at risk for deficiency (4,23).

Table 3. U.S. RDA for Vitamin B₆

Age	RDA, mg
infants and children <4 yr	0.7
adults and children ≥4 yr	2.0
pregnant or lactating women	2.5

The typical U.S. daily diet contains 1.1–3.6 mg of vitamin B₆, most coming from meats and vegetables. Poor diets may provide less than half of these amounts and less than the RDA. Some populations require higher amounts: persons with high protein intakes, pregnant and lactating women, users of oral contraceptives, alcoholics, users of drugs which interfere with vitamin B₆ function, and those afflicted with some diseases. Several reviews have examined the relationship of vitamin B₆ and specific diseases in more detail (4,23).

Manufacture

Pyridoxal (2) can be prepared by oxidation of pyridoxine (1) (13). Pyridoxamine (3) can be made by reduction of pyridoxal oxime (32). Only pyridoxine (1), the most stable of the three, is manufactured. The earliest syntheses, by degradation of isoquinoline rings, are only of historical interest. The first syntheses to be applied on a large scale relied on classical condensation reactions. Many variations were explored during the 1940s and 1950s and some formed the basis of early production. Such routes required numerous steps giving low overall yields, in part due to difficulties of obtaining the correct oxidation levels of the 4- and 5-substituents. These methods have been reviewed (11,12). In the mid-1950s, electrooxidation of furan derivatives was shown to provide direct precursors to hydroxypyridines (33). In later improvements, cycloaddition of readily available 2-butyne-1,4-diol to 4-methyloxazole [693-93-6] (11) followed by amidoalkylation conveniently provided the furans (34,35). The overall yields of these linear sequences are modest (Fig. 4).

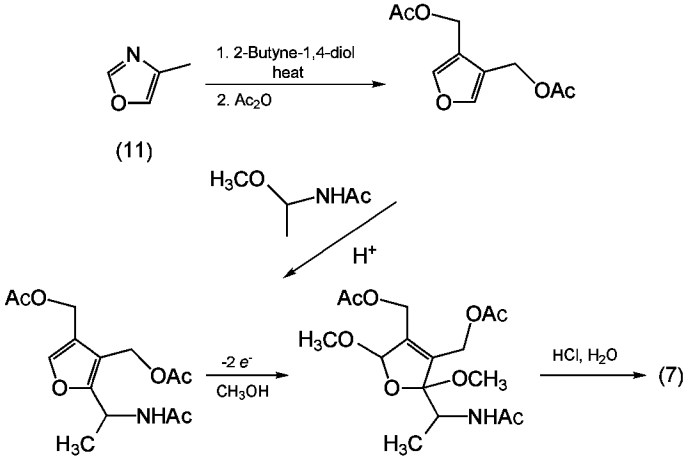


Fig. 4. Furan route to pyridoxine.

The current paradigm for B₆ syntheses came from the first report in 1957 of a synthesis of pyridines by cycloaddition reactions of oxazoles (36) (Fig. 5). This was adapted for production of pyridoxine shortly thereafter. Intensive research by Ajinomoto, BASF, Daiichi, Merck, Roche, Takeda, and other companies has resulted in numerous publications and patents describing variations. These routes are convergent, shorter, and of reasonably high throughput.

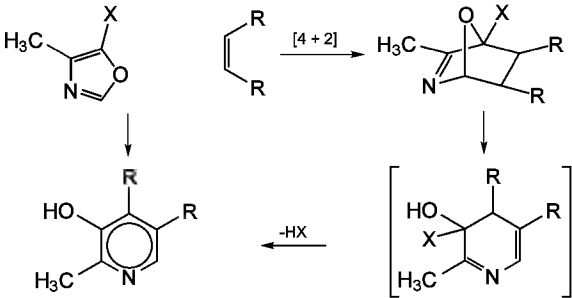


Fig. 5. Synthesis of pyridines by cycloaddition reactions of oxazoles.

Substitution on the azadiene and alkene affect reactivity and only certain pairs react efficiently. Electrophilic alkenes and oxazoles bearing donor groups generally allow faster reactions at lower temperatures (37,38). Steric effects play a lesser role. Earlier variations with 4-methyloxazole (11) and maleic acid derivatives require the presence of oxidants such as hydrogen peroxide or nitroarenes to aromatize the cycloadducts in order to obtain good yields (11). Subsequent costly reduction steps are also needed to correct the oxidation level of the product side chains. One advantage is the use of leaving groups on one of the cycloaddition partners to facilitate aromatization without loss of the product 3-hydroxyl group. These superfluous groups can be added at the 5-position of the oxazoles or on the alkene.

Oxazoles preferred in practice bear 5-ethoxy- or 5-cyanosubstituents. Other alkoxy (39–42), trimethylsilyloxy (43), alkylthio (44), amino (45), and

carboalkoxy (46) substituents are reported. Processes using 5-ethoxyoxazoles (12), (13), and (14) have been published (38,47–60). Extra carbons carried by the Takeda, Chinese, and Ajinomoto intermediates are lost as waste by decarboxylation. A process with 5-cyano-4-methyloxazole (15) has been disclosed by Roche (59). Reactions of these azadienes with butenediol derivatives, the usual alkene partner, take place at higher temperatures (115–180°C) and require longer reaction times (10–20 h) than those with maleate partners. The ethoxy compound (12) reacts under milder conditions and the adduct is isolable. Mild acid converts it to the product and ethanol. With the higher temperatures required of the cyano compound [1003-52-7] (15), the intermediate cycloadduct is converted directly to the product by elimination of waste hydrogen cyanide. Often the reactions are run with neat liquid reagents having an excess of alkene as solvent. Polar solvents such as sulfolane and *N*-methyl-pyrrolidone are claimed to be superior for reactions of the ethoxy compound with butenediol (53). Organic acids, phenols, maleic acid derivatives, and inorganic bases are suggested as catalysts (51,52,54,59,61,62) (Fig. 6).

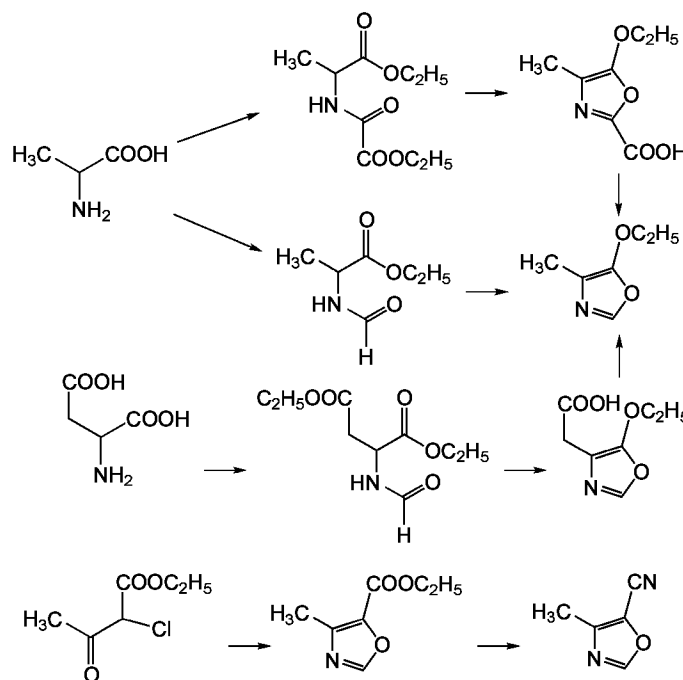


Fig. 6. Process chemistry for oxazoles.

As the alkene partners, (*Z*)-2-butene-1,4-diol acetals, ketals, esters or ethers (11), and a polymer-bound acetal (63) are described. Butenediol may be used if polar solvents are employed (54). Current manufacturers generally use the cyclic acetal [5417-35-6] (16) of butenediol with isobutyraldehyde, a liquid of conveniently high boiling point that readily allows the product to be deprotected with no waste (Fig. 7). Butenediol is readily and cheaply available from the reaction of acetylene and formaldehyde to 2-butyne-1,4-diol, then catalytic hydrogenation. In an unusual variant of this cycloaddition approach, a leaving group on the alkene facilitates aromatization. Cyclic sulfone [41409-84-1] (17) paired with 4-methyloxazole (11) effectively gives the pyridine nucleus (64,65) taking advantage of the increased reactivity of an electrophilic alkene while maintaining the correct oxidation level of the side chains. Electrophilic dihydrofuran [332-77-4] (18) reacts efficiently with excess ethoxymethyloxazole (12), also at lower temperatures (66).

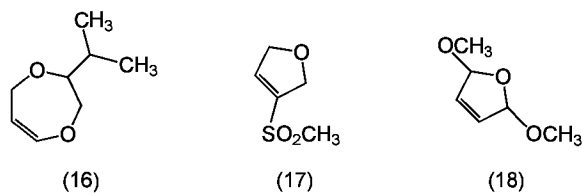


Fig. 7. Dienophiles.

Ethoxyoxazole (12) is constructed from *N*-formyl alanine esters (67) and ethoxyoxazole acid (13) from *N*-oxalyl alanine esters (55–58). Alanine is readily available from China by chlorination of propionic acid, then ammonation. Both ethoxyoxazole (12) and the acid (14) are derived from maleic anhydride by *N*-formyl aspartate esters (52,68). Cyanooxazole (15) is derived from ketene by ethyl chloroacetoacetate and formamide (69) or, alternatively, from diketene (70). All routes require dehydration of amides as a critical step. Traditionally these dehydrations have been accomplished with phosphorus pentoxide, phosphoryl chloride, or phosgene; however, more environmentally acceptable procedures use acetic anhydride (71), cyanuryl chloride (72), or gas-phase catalytic dehydration (73). In some cases, the oxazoles are generated *in situ* in the cycloaddition reactions from *N*-formyl compounds by the isonitriles under dehydrating conditions (45,74,75). Other cycloaddition routes include reactions of 1,2,4-triazines with butynediol or enamine derivatives give the pyridoxine nucleus (76,77), or a cobalt-catalyzed [2 + 2 + 2] cycloaddition of acetylenic ethers with acetonitrile (78,79) (Fig. 8).

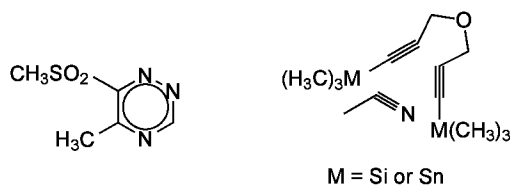


Fig. 8. Other cycloaddition routes.

Economic Aspects

Significant producers include Daiichi, Hoffmann-La Roche, Takeda, and several factories in mainland China. Takeda and Daiichi practice processes based on alanine by ethoxyoxazole (12), at least one Chinese producer from alanine by oxazole acid (13), and Roche from ketene via cyanooxazole (15).

Production is practiced at many hundreds of metric tons annually by batchwise or semicontinuous operation in automated glass and stainless steel equipment typical of fine chemical manufacture. World production in 1995 was estimated at about 3000 MT with sales prices in the United States in the range of \$22–\$30 per kilogram.

Product Specifications and Testing

Nutrients and diet supplements without claims of therapeutic effects are considered foods, and are thus regulated by the U.S. Food and Drug Administration. These are further subject to specific food regulations. Specifications for pyridoxine hydrochloride (7) for foods are given in the *Food Chemicals Codex* (80) and for pharmaceuticals in the *U.S. Pharmacopeia* (81). General test methods have been summarized (82).

Safety and Handling

Pyridoxine hydrochloride is typically packaged in standard 20–25 kg polylined fiber cartons or drums. Smaller packing sizes are also available. Pyridoxine hydrochloride is kept under normal dry storage conditions not to exceed 25°C and protected from light. It is listed in the TSCA inventory and material safety data sheets are available from suppliers. Protection against dust exposure to the eyes, skin, or lungs (IOEL of 2.0 mg m³ time weighted average) and fire and dust explosion are indicated. NPFA ratings for health, fire, and reactivity are 1, 2, and 1. Neither national nor international transportation is regulated. The U.S. EPA has not established reportable quantities for environmental releases. The acute LD₅₀ (rat oral) of 7750 mg kg classifies this material as practically nontoxic orally (83). Oral doses of 100–300 mg d (50–150 times the RDA) have been used in therapeutic studies without adverse effects. Excess amounts are readily cleared by the kidney. More recent studies show that long-term consumption of very large doses can lead to sensory nerve damage. Supplements of more than 200 mg d should be taken only under medical supervision. Intravenous administration is not associated with toxicity (84).

Analytical Methods

Determining vitamin B in foods and tissues is complicated by the number of forms, typical low levels, varying degrees of ionization, lability to heat, light, and base, interconversion, and binding to proteins and glucose. Colorimetric, fluorometric, microbiological, animal, and chromatographic methods are used and have been reviewed in depth (5,11,85,86). Few are fully suitable for simultaneous analysis of all six bioactive forms at the typical low levels. Colorimetric assays involve derivatization, are relatively insensitive, and used only for assay of high potency samples such as premixes. Microbiological assays have previously been the most commonly used, especially for foods. The yeast *Saccharomyces uvarum* (ATCC 9080, also known as *S. carlsbergensis* 4228) responds well to pyridoxine and pyridoxal, less to pyridoxamine, and not at all to the phosphate forms. Hydrolysis is first necessary; by the official AOAC method, the matrix is autoclaved with hydrochloric acid (87). Such methods are slow and overestimate bioavailability. Animal assays in vitamin B -deficient chicks or rats are time consuming and expensive but have been useful in determining bioavailability of derivatives and related materials.

Increasingly, chromatographic methods are used for their ease, accuracy, and ability to determine all forms simultaneously. Gas–liquid chromatography is hampered by the polarity of the compounds unless silyl or acyl derivatives are used (5,88). The comprehensive method of choice generally is high pressure liquid chromatography (hplc) (5). Separations by ion exchange or reverse-phase absorption with ion-pair reagents and isocratic or gradient mobile phases are common. Direct uv detection is less sensitive but useful for premixes and vitamin mixtures. Post-column derivatization and fluorescence detection allows measurement to nanogram levels. High performance capillary zone electrophoresis (cze) coupled with fluorescence detection allows measurement to below picogram levels (89). Gc ms, hplc ms, and cze ms hold promise for measurement and identification of femtomole levels of vitamers and metabolites (90–92). Immunoassay (qv) method are being developed (4).

Biosynthesis

Almost all microorganisms synthesize B vitamers intracellularly at levels of 0.05–0.3 mg g dry weight of cells and excrete at levels of 0.05–0.5 mg L. Biosynthesis in *E. coli* is known to be controlled by both feedback inhibition and repression. Some organisms can overproduce, for example *B. subtilis*, up to 2.0–5.0 mg L, *Flavobacterium* sp. 238-7 up to 20 mg L, and yeast *Pichia guilliermondii* up to 25 mg L (4,93). The mechanisms of vitamin B biosynthesis are partly understood. Extensive isotopic label feeding studies with *E. coli* mutants show all carbons are derived from glucose and the nitrogen from glycine. Union of the intermediates 1-deoxy-D-xylulose [16709-34-9] (19) and 4-hydroxy-L-threonine [21768-45-6] (20) gives pyridoxine in several steps, only two of which require enzymatic catalysis (Fig. 9) (94). In aerobic bacteria, pyridoxal phosphate is made as needed by the action of oxidase or dehydrogenase enzymes on pyridoxine phosphate or pyridoxamine phosphate (4,93). The pathways in *Flavobacterium* sp 238-7 (95) and in *Saccharomyces cerevisiae* (96) may differ from that in *E. coli* in the early steps. In microorganisms, some genes and related enzymes for biosynthesis have been identified (94,97–99). Significant additional progress will be necessary before cost-effective production of vitamin B by bioprocesses is possible.

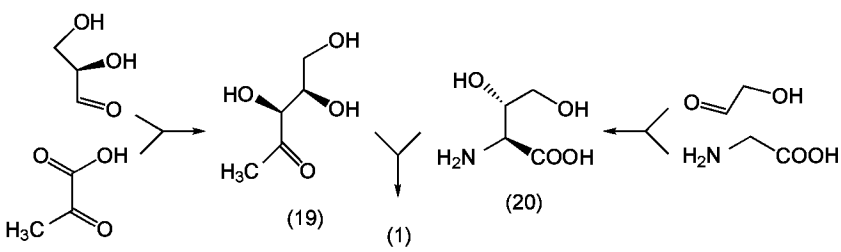


Fig. 9. Biosynthetic pathway to pyridoxine.

Analogues and Antagonists

Over 250 analogues of the B vitamers have been reported (11,100). Nearly all have low vitamin B activity and some show antagonism. Among these are the 4-deshydroxy analogue, pyridoxine 4-ethers, and 4-amino-5-hydroxymethyl-2-methylpyrimidine, a biosynthetic precursor to thiamine. Structurally unrelated antagonists include drugs such as isoniazid, cycloserine, and penicillamine, which are known to bind to pyridoxal enzyme active sites (4).

Uses

The primary uses of pyridoxine hydrochloride are in multivitamin supplement tablets and for fortification of human food and animal feed, especially for poultry and pigs. Most breakfast cereals and infant formulas in the United States are supplemented. Lesser amounts are used therapeutically to correct deficiencies or to treat specific disorders. Pyridoxine hydrochloride has been used experimentally to treat a variety of conditions with varying degrees of effectiveness (4,23). Pyridoxine hydrochloride is readily incorporated into premixes and foods.

BIBLIOGRAPHY

"Vitamins, Pyridoxine" in *ECT* 1st ed., Vol. 11, pp. 293–307, by S. A. Harris, G. E. Sita, and P. G. Stecher, Merck & Co.; in *ECT* 2nd ed., Vol. 16, pp. 806–824, by S. A. Harris, E. E. Harris, and R. W. Burg, Merck & Co.; in *ECT* 3rd ed., Vol. 24, pp. 94–107, by D. L. Coffen, Hoffmann-La Roche Inc.

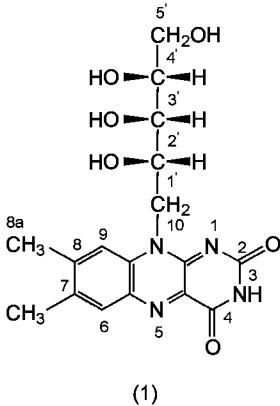
- Commission on Biological Nomenclature, *Biochem. J.* **137**, 417 (1974).
- S. A. Harris and K. Folkers, *J. Am. Chem. Soc.* **61**, 1245 (1939).
- R. Kuhn and co-workers, *Naturwissenschaften* **27**, 469 (1939).
- D. Dolphin, R. Poulson, and O. Avramovic, eds., *Coenzymes and Cofactors*, Vols. 1A & 1B, John Wiley & Sons, Inc., New York, 1986.
- J. Ubbink, in A. P. De Leenheer, W. Lambert, and H. Nelis, eds., *Modern Chromatographic Analysis of Vitamins*, 2nd ed., Marcel Dekker, New York, 1992, Chapt. 10.
- W. Korytnyk and R. P. Singh, *J. Am. Chem. Soc.* **85**, 2813 (1963).
- I. D. Spenser and co-workers, *J. Biol. Chem.* **262**, 7463 (1987).
- A. Lycka and J. Cizmarik, *Pharmazie* **45**, 371 (1990).
- W. Korytnyk and co-workers, *J. Am. Chem. Soc.* **88**, 1233 (1966).
- J. Iida and T. Murata, *Anal. Sci.* **6**, 273 (1990).
- O. Isler, G. Brubacher, S. Ghisla, and B. Kraeutler, eds., *Vitamine II*, Verlag, New York, 1981, pp. 193–227.
- J. M. Osbond, *Vitamins and Hormones* **22**, 387 (1964).
- K. Folkers and co-workers, *J. Am. Chem. Soc.* **73**, 3430 (1951) and cites therein.
- S. A. Harris, *J. Am. Chem. Soc.* **63**, 3363 (1941).
- Nutr. Rev.* **36**, 346 (1978).
- The Merck Index*, 11th ed., Merck & Co., Rahway, N.J., 1989, p. 1269.
- G. S. Shephard and D. Labadarios, *Clin. Chim. Acta* **60**, 307 (1981).
- B. Saida and J. Warthesen, *J. Agric. Food Chem.* **31**, 876 (1983).
- C. Y. W. Ang, *J. Assoc. Official Anal. Chem.* **62**, 1170 (1979).
- A. R. Forester and I. G. Davidson, *Radiat. Phys. Chem.* **36**, 403 (1990).
- K. Tadera and co-workers, *Agric. Biol. Chem.* **52**, 2359 (1988).
- M. L. Hu and co-workers, *Chem. Biol. Interact.* **97**, 63 (1995).
- J. E. Leklem, in L. J. Machlin, ed., *Handbook of the Vitamins*, 2nd ed., Chapt. 6, Marcel Dekker, New York, 1991.
- M. L. Orr, *Pantothenic Acid, Vitamin B₆ and Vitamin B₁₂ in Foods*, Home Economics Report No. 36, U.S. Dept. of Agriculture, Washington, D.C., 1969.
- J. F. Gregory and D. B. Sartain, *J. Agr. Food Chem.* **39**, 899 (1991).
- K. Tadera and Y. Naka, *Agr. Biol. Chem.* **55**, 563 (1991).
- K. D. Lunan and C. A. West, *Arch. Biochem. Biophys.* **101**, 262 (1963).
- J. Gregory and co-workers, *J. Food Sci.* **51**, 1345 (1986).
- J. B. Tarr, T. Tamura, and E. L. R. Stokstad, *Am. J. Clin. Nutr.* **34**, 1328 (1981).
- A. H. Merrill and M. J. Henderson, *Ann. N.Y. Acad. Sci.* **585**, 110 (1990).
- Committee on Dietary Allowances, Food and Nutrition Board, *Recommended Daily Allowances*, 10th ed., National Academy of Sciences, Washington, D.C., 1980, pp. 142–150.
- P. Karrer and co-workers, *Helv. Chim. Acta* **32**, 1004 (1948).
- N. Clauson-Kaas and co-workers, *Acta Chem. Scand.* **9**, 1,9,14,17,23 (1955).
- Ger. Pat. 1,935,009** (Jan. 14, 1971), F. Graf and H. Konig (to BASF).
- Y. Matsumura and co-workers, *Chem. Lett.*, 1121 (1981).
- M. Y. Karpeiskii and V. L. Flornet'ev, *Russ. Chem. Rev.* **38**, 540 (1969).
- I. Turchi and M. J. S. Dewar, *Chem. Rev.* **75**, 388 (1975).
- G. Y. Kondrat'eva and co-workers, *Dokl. Akad. Nauk SSSR* **142**, 593 (1962).
- T. Naito and co-workers, *Chem. Pharm. Bull.* **13**, 873 (1965).
- Z. I. Itov and co-workers, *Khim. Farm. Zh.* **13**, 59 (1980).
- E. E. Harris and co-workers, *Tetrahedron* **23**, 943 (1967).
- Ger. Pat. 3,029,231** (Nov. 3, 1982), H. R. Mueller (to Merck).
- H. Takagaki and co-workers, *Chem. Lett.*, 183 (1979).
- Jpn. Pat. 69-23,088** (Oct. 1, 1969), T. Naito, K. Ueno, and T. Miki (to Daiichi Seiyaku Co, Ltd.).
- Jpn. Pat. 81-128,761** (Oct. 8, 1981) (to Daiichi Seiyaku Co, Ltd.).
- H. Murakami and M. Iwanami, *Bull. Chem. Soc. Jpn.* **41**, 726 (1968).
- E. E. Harris and co-workers, *J. Org. Chem.* **27**, 2705 (1962).
- U.S. Pat. 3,227,721** (May 24, 1965), E. E. Harris, K. Pfister, and R. A. Firestone (to Merck & Co.).
- Jpn. Pat. 73-30,636** (Sept. 21, 1973), Y. Morita, S. Onishi, and T. Yamagami (to Daiichi Seiyaku Co.).
- Jpn. Pat. 70-39,259** (Dec. 10, 1970), T. Naito, Y. Morito, and U. Toshihiro (to Daiichi Seiyaku Co.).
- Jpn. Pat. 72-39,115** (Oct. 3, 1972), T. Miki and T. Matsuo (to Takeda Chemical Industries, Ltd.).
- Jpn. Pat. 81-99,461** (Aug. 10, 1981) (to Takeda Chemical Industries, Ltd.).
- T. Miki and T. Matsuo, *Yakugaku Zasshi*, **87**, 323 (1967).
- U.S. Pat. 3,796,720** (Mar. 12, 1974), H. Miki and H. Saikawa (to Takeda Chemical Industries).
- J. Maeda and co-workers, *Bull. Chem. Soc. Jpn.* **42**, 1435 (1969).
- Fr. Pat. 1,533,817 (July 19, 1968) (to Ajinomoto Co.).
- H. Zhou and co-workers, *Zhongguo Yiyao Gongye Zazhi* **25**, 385 (1994).
- China. Pat. 86-101,512 (July 7, 1986), H. Zhou (to Shanghai Industrial Res. Inst.).
- U.S. Pat. 3,250,778** (Nov. 11, 1962), W. Kimel and W. Leimgruber (to Hoffmann-La Roche Inc.).
- M. Dunic and co-workers, *Prax. Vet.* **37**, 81 (1989).
- D. Szlopmek-Nesteruk and co-workers, *Przem. Chem.* **54**, 238 (1975).
- U.S. Pat. 3,822,274** (July 2, 1974), E. E. Harris and R. Currie (to Merck & Co.).
- H. Ritter and R. Sperber, *Macromolecules* **27**, 5919 (1994).
- W. Boell and H. Koenig, *Ann. Chem.*, 1657 (1979).
- U.S. Pat. 3,876,649** (Apr. 8, 1975), W. Boell and H. Koenig (to BASF).
- T. Naito and co-workers, *Tetr. Lett.*, 5767 (1968).
- P. Karrer and C. Granacher, *Helv. Chem. Acta* **7**, 763 (1924).
- Jpn. Pat. 68-30,286** (Dec. 26, 1968), T. Kitazawa, T. Matsuyama, K. Kunimitsu, and K. Masuya (to Fujisawa Pharmaceutical Co.).
- T. Rinderspacher and B. Prijs, *Helv. Chim. Acta* **43**, 1522 (1960).
- U.S. Pat. 4,026,901** (June 6, 1978), D. L. Coffen (to Hoffmann-La Roche Inc.).

71. U.S. Pat. **4,011,234** (Mar. 8, 1977), H. Hoffmann-Paquotte (to Hoffmann-La Roche Ltd.).
72. Eur. Pat. **612736** (Aug. 31, 1994), W. Bonrath, R. Karge, and H. Pauling (to Hoffmann-La Roche Ltd.).
73. Eur. Pat. **492,233** (July 1, 1992), P. Noesberger (to Hoffmann-La Roche Ltd.).
74. S. Shimada and M. Oki, *Chem. Pharm. Bull.* **32**, 38 (1984).
75. Jpn. Pat. **79-20,493** (July 23, 1979), S. Ohnishi and Y. Morita (to Daiichi Seiyaku Co.).
76. M. Oki and S. Shimada, *Chem. Pharm. Bull.* **35**, 4705 (1987).
77. E. C. Taylor and J. Macor, *J. Org. Chem.* **54**, 1249 (1989).
78. R. E. Geiger and co-workers, *Helv. Chim. Acta* **67**, 1274 (1984).
79. C. A. Parnell and K. P. C. Vollhardt, *Tetr.* **41**, 5791 (1985).
80. Food and Nutrition Board, National Research Council, *Food Chemicals Codex* 3rd ed., National Academy Press, Washington, D.C., 1981, p. 260.
81. United States Pharmacopieal Convention, Inc., *The United States Pharmacopeia XXII* (USP XXII-NF XVII), Rockeville, Md., 1990, p. 1194.
82. H. Y. Aboul-Enein and M. A. Loutfy, in K. Florey, ed., *Analytical Profiles of Drug Substances*, Vol. 13, Academic Press, New York, 1984, p. 447.
83. *Pyridoxine Hydrochloride*, MSDS 11076, Hoffmann-LaRoche Inc., Nutley, N.J., 1997.
84. M. Cohen and A. Bendich, *Toxicol. Lett.* **34**, 129 (1986).
85. M. M. Polansky, in J. Augustin, B. P. Klein, D. A. Decker, and P. B. Venugopal, eds., *Methods of Vitamin Assay*, 4th ed., Association of Vitamin Chemists, John Wiley & Sons, Inc., New York, 1985, Chapt. 17.
86. J. F. Gregory, *J. Food Comp. Anal.* **1**, 105 (1988).
87. K. Helrich, ed., *Official Methods of Analysis of the Association of Official Analytical Chemists*, 15th ed., Association of Official Analytical Chemists, Inc., Arlington, Va., 1990, p. 961.15.
88. S. P. Colburn and J. D. Mahuren, *J. Biol. Chem.* **262**, 2642 (1987).
89. Y. Kurosu and co-workers, *Bitamin* **66**, 91 (1992).
90. R. D. Smith and co-workers, *Anal. Chem.* **60**, 436 (1988).
91. R. Benyon and co-workers, *J. Chrom.* **581**, 179 (1992).
92. J. Iida and T. Murata, *Anal. Sci.* **6**, 273 (1990).
93. Y. Tani, in E. Vandamme, ed., *Biotechnology of Vitamins, Pigments and Growth Factors*, Elsevier, New York, 1988, Chapt. 13.
94. R. E. Hill, I. D. Spenser, and co-workers, *J. Am. Chem. Soc.* **117**, 1661 (1995).
95. M. Morisaki, *Kurume Igakkai Zasshi* **53**, 649 (1990).
96. K. Tazuya and co-workers, *Biochem. Biophys. Acta*, **1244**, 113 (1995).
97. M. Winkler and co-workers, *J. Bacteriol.* **177**, 2804 (1995) and cites therein.
98. M. Winkler and co-workers, *FEMS Microbiol. Lett.* **135**, 275 (1996).
99. T. Copeland and co-workers, *J. Bacteriol.* **174**, 1544 (1992).
100. W. Korytnik and M. Ikawa, in S. P. Colowick and N. O. Kaplan, eds., *Methods in Enzymology*, Vol. 18A, Academic Press, New York, pp. 524–566.

David Burdick
Hoffmann-La Roche Inc.

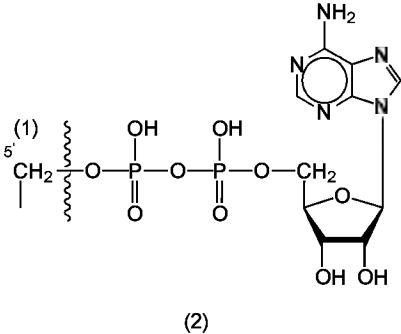
RIBOFLAVIN (B₂)

Riboflavin [83-88-5] (vitamin B₂, vitamin G, lactoflavin, ovoflavin, lyochrome, hepatoflavin, uroflavin) has the chemical name 1-deoxy-1-(3,4-dihydro-7,8dimethyl-2,4-dioxobenzo[g]pteridin-10(2*H*)-yl)-D-ribitol, 7,8-dimethyl-10-D-ribitylisoalloxazine, C₁₇H₂₀N₄O (1), mol wt 376.37.



In 1933, R. Kuhn and his co-workers first isolated riboflavin from eggs in a pure, crystalline state (1), named it ovoflavin, and determined its function as a vitamin (2). At the same time, impure crystalline preparations of riboflavin were isolated from whey and named lyochrome and, later, lactoflavin. Soon thereafter, P. Karrer and his co-workers isolated riboflavin from a wide variety of animal organs and vegetable sources and named it hepatoflavin (3). Ovoflavin from egg, lactoflavin from milk, and hepatoflavin from liver were all subsequently identified as riboflavin. The discovery of the yellow enzyme by Warburg and Christian in 1932 and their description of lumiflavin (4), a photochemical degradation product of riboflavin, were of great use for the elucidation of the chemical structure of riboflavin by Kuhn and his co-workers (5). The structure was confirmed in 1935 by the synthesis by Karrer and his co-workers (6), and Kuhn and his co-workers (7).

For therapeutic use, riboflavin is produced by chemical synthesis, whereas concentrates for poultry and livestock feeds are manufactured by fermentation using microorganisms such as *Ashbya gossypii* and *Eremothecium ashbyii*, which have the capacity to synthesize large quantities of riboflavin. In the free form, riboflavin occurs in the retina of the eye, in whey, and in urine. Principally, however, riboflavin fulfills its metabolic function in a complex form. In general, riboflavin is converted into flavin mononucleotide (FMN, riboflavin-5'-phosphate) and flavin-adenine dinucleotide (FAD) (2), which serve as the prosthetic groups (coenzymes), ie, they combine with specific proteins (apoenzymes) to form flavoenzymes, in a series of oxidation–reduction catalysts widely distributed in nature. In several riboflavin coenzymes, the apoenzyme is covalently attached to C^{8a} through a linkage to the nitrogen of a histidine imidazolyl group, to the sulfur of a cysteine residue, or to the oxygen of a tyrosine moiety (see AMINO ACIDS). Riboflavin is not a nucleotide, since it is derived from D-ribitol rather than D-ribose, and therefore FMN and FAD are not truly nucleotides; yet this designation has been accepted overwhelmingly and continues to be used.



As a coenzyme component in tissue oxidation–reduction and respiration, riboflavin is distributed in some degree in virtually all naturally occurring foods. Liver, heart, kidney, milk, eggs, lean meats, malted barley, and fresh leafy vegetables are particularly good sources of riboflavin (see Table 1). It does not seem to have long stability in food products (8).

Table 1. Riboflavin Content of Various Foods^a

Food	mg/100 g	Food	mg/100 g
fruits		fish	
apple, raw	0.01	cod, haddock, raw	0.17
banana, raw	0.04	salmon, raw	0.17
citrus, grapefruit, orange	0.03–0.04	salmon, canned	0.12
strawberry	0.03	tuna, canned	0.13
vegetables		whitefish, herring, halibut	0.17–0.29
broccoli, raw	0.27	grain	
cabbage, raw	0.05	corn, entire	0.10
fresh green peas	0.14	wheat, entire	0.10
mushroom	0.57	wheat, germ	0.6
parsley	0.24	rice, entire	0.06
potato, raw	0.03	rye, entire	0.20

sweet corn	0.14	cereal products	
sweet potato, raw	0.05	refined	
tomato, raw	0.03	bread	0.07–0.10
meat		cereal	0.10–0.15
beef muscle	0.16–0.32	soda cracker	0.02–0.10
pork muscle	0.19–0.33	whole grain and enriched	
chicken muscle	0.10–0.31	bread	0.12–0.25
liver, beef, pork	3.00–3.60	cereal	0.20–1.25
		dairy products	
		cheese	0.33–0.68
		eggs	0.48
		milk	0.15–0.18

^a Averages from several sources, often encompassing a wide range of analytical results; should be regarded as working estimates which vary with geography, season, and preparative method.

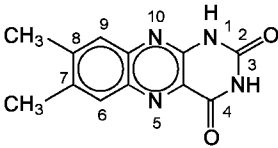
Riboflavin is widely used in the pharmaceutical, food-enrichment, and feed-supplement industries. Riboflavin USP is administered orally in tablets or by injection as an aqueous solution, which may contain nicotinamide or other solubilizers. As a supplement to animal feeds, riboflavin is usually added at concentrations of 2–8 mg kg, depending on the species and age of the animal (see FEEDS AND FEED ADDITIVES).

Properties

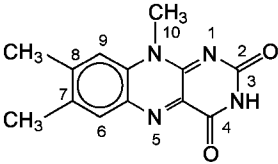
Riboflavin forms fine yellow to orange-yellow needles with a bitter taste from 2 N acetic acid, alcohol, water, or pyridine. It melts with decomposition at 278–279°C (darkens at ca 240°C). The solubility of riboflavin in water is 10–13 mg 100 mL at 25–27.5°C, and in absolute ethanol 4.5 mg 100 mL at 27.5°C; it is slightly soluble in amyl alcohol, cyclohexanol, benzyl alcohol, amyl acetate, and phenol, but insoluble in ether, chloroform, acetone, and benzene. It is very soluble in dilute alkali, but these solutions are unstable. Various polymorphic crystalline forms of riboflavin exhibit variations in physical properties. In aqueous nicotinamide solution at pH 5, solubility increases from 0.1 to 2.5% as the nicotinamide concentration increases from 5 to 50% (9).

In aqueous solution, riboflavin has absorption at ca 220–225, 226, 371, 444 and 475 nm. Neutral aqueous solutions of riboflavin have a greenish yellow color and an intense yellowish green fluorescence with a maximum at ca 530 nm and a quantum yield of Φ_f = 0.25 at pH 2.6 (10). Fluorescence disappears upon the addition of acid or alkali. The fluorescence is used in quantitative determinations. The optical activity of riboflavin in neutral and acid solutions is [α]_D²⁰ +56.5–59.5° (0.5% , dil HCl). In an alkaline solution, it depends upon the concentration, eg, [α]_D²⁶ –112–122° (50 mg in 2 mL 0.1 N alcoholic NaOH diluted to 10 mL with water). Borate-containing solutions are strongly dextrorotatory, because borate complexes with the ribityl side chain of riboflavin; [α]_D²⁰ +340° (pH 12).

Photochemical decomposition of riboflavin in neutral or acid solution gives lumichrome (3), 7,8-dimethylalloxazine, which was synthesized and characterized by Karrer and his co-workers in 1934 (11). In alkaline solution, the irradiation product is lumiflavin (4), 7,8,10-trimethylisoalloxazine; its uv–vis absorption spectrum resembles that of riboflavin. It was prepared and characterized in 1933 (5). Another photodecomposition product of riboflavin is 7,8-dimethyl-10-formylmethylisoalloxazine (12).

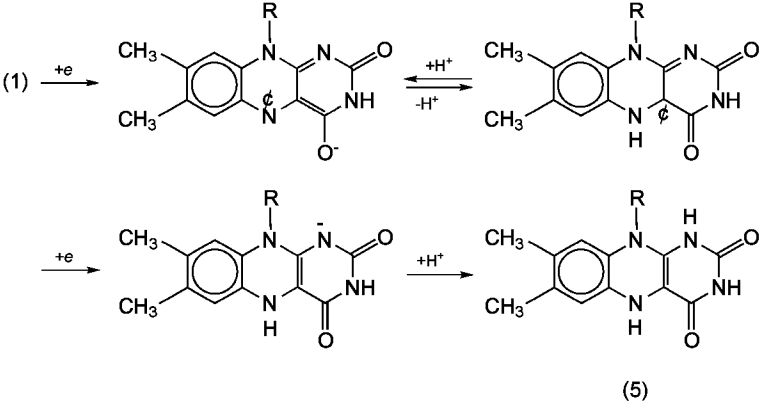


(3)



(4)

Riboflavin is stable against acids, air, and common oxidizing agents such as bromine and nitrous acid (except chromic acid, KMnO₄, and potassium persulfate). Upon reduction by conventional agents such as sodium dithionite, Na₂S₂O₄, zinc in acidic solution, or catalytically activated hydrogen, riboflavin readily takes up two hydrogen atoms to form the almost colorless 1,5-dihydropyrimidin-2-one (5) (Fig. 1), which is reoxidized by shaking with air. This oxidation–reduction system has considerable stability, a normal potential of –0.208 V (referred to as the normal hydrogen electrode), and is probably responsible for the physiological functions of riboflavin. The flavins are reduced to dihydroflavins through intermediate semiquinone radicals (13,14), which have been directly observed by electron spin resonance (esr) (15–18) and electron double resonance (ENDOR) (19).



R = D-Ribityl

Fig. 1. Formation of dihydronitroflavin.

Riboflavin forms a deep-red silver salt (1). The strong bathochromic shift of the spectra of riboflavin analogues occurring by interaction with Ag^+ can also be obtained with Cu^+ and Hg^{2+} complexes (20). These complexes contain the flavin and the metal ligand anion in a ratio of 1:1. Their color is the result of a charge transfer between the metal and flavin (21). The chelates with Fe(II-III) , Mo(V-VI) , Cu(I-II) , and Ag(I-II) belong to this group; the last two are stable in the presence of water. Another group of metal complexes, radical chelates, are formed with Mn(II) , Fe(II) , Co(II) , Ni(II) , Zn(II) , and Cd(II) ; in these cases, the radical character of the ligand is conserved (22).

Chemical Synthesis

In 1935, Karrer (6) and Kuhn (7) each proved independently that riboflavin was 7,8-dimethyl-10-D-ribitylisoalloxazine by total synthesis (see Fig. 2). These syntheses are essentially the same and involve a condensation of 6-D-ribitylamino-3,4-xylydine (6) with alloxan (7) in acid solution. Boric acid as a catalyst increases the yield considerably (23). The intermediate (6) was prepared by a condensation of 6-nitro-3,4-xylydine (8) with D-ribose (9), followed by catalytic reduction of the riboside (10). The yield based on D-ribose was increased (24) by using N-D-ribityl-3,4-xylydine (11), which was prepared by the condensation of 3,4-xylydine (12) with D-ribose (9), followed by catalytic reduction. The reduced product (11) was coupled with *p*-nitrophenyldiazonium salt to give 4,5-dimethyl-2-*p*-nitrophenylazo-1-D-ribitylamino benzene (13), which was reduced to (6) and treated with alloxan (7) (25) to give riboflavin. Replacement of (7) by 5,5-dichlorobarbituric acid (26), 5,5-dibromobarbituric acid, or 5-bromobarbituric acid (27) in the above syntheses yields riboflavin.

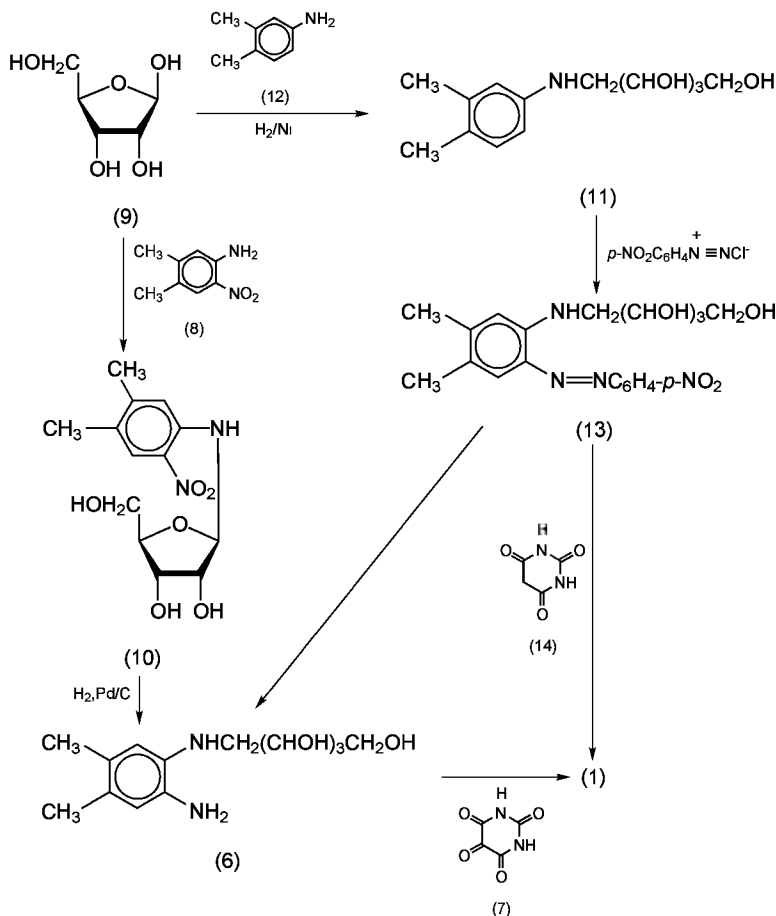


Fig. 2. Syntheses of riboflavin.

More conveniently, compound (13) was directly condensed with barbituric acid (14) in acetic acid (28) or in the presence of an acid catalyst in an organic solvent (29). The same azo dye intermediate (13) and alloxantin give riboflavin in the presence of palladium on charcoal in alcoholic hydrochloric acid under nitrogen. This reaction may involve the reduction of the azo group to the *o*-phenylenediamine by the alloxantin, which is dehydrogenated to alloxan (see UREA) (30).

Although it is not suitable for large-scale manufacture, the synthesis of riboflavin from lumazine derivatives is interesting in connection with the biosynthesis of riboflavin (see Fig. 3). Thus, 5-amino-6-D-ribitylaminouracil (15) was condensed with a dimeric or trimeric aldol of biacetyl to give riboflavin (1) through the formation of intermediary 6,7-dimethyl-8-D-ribityllumazine (16) (31). A variation of the above synthesis involves the condensation of monomeric biacetyl and preformed (16) prepared by the condensation of (15) with biacetyl (32). The condensation of (15) with 4,5-dimethyl-1,2-benzoquinone (17) is another pathway to riboflavin, although in low yield (33).

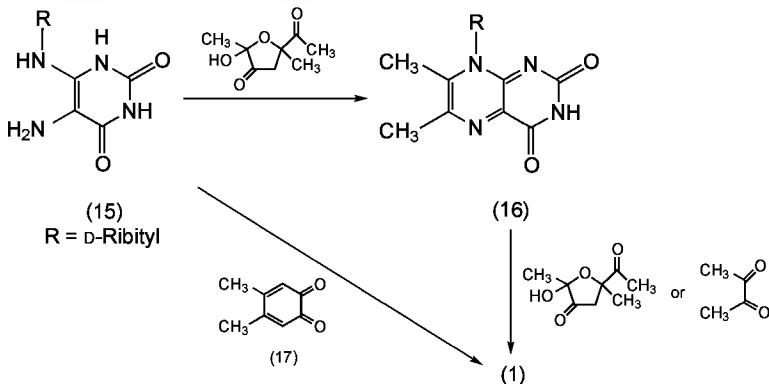


Fig. 3. Syntheses of riboflavin from a lumazine.

Later, a completely different and more convenient synthesis of riboflavin and analogues was developed (34). It consists of the nitrosative cyclization of 6-(*N*-D-ribityl-3,4-xylydino)uracil (18), obtained from the condensation of *N*-D-ribityl-3,4-xylydine (11) and 6-chlorouracil (19), with excess sodium nitrite in acetic acid, to give riboflavin-5-oxide (20) in high yield. Reduction with sodium dithionite gives (1). In another synthesis, 5-nitro-6-(*N*-D-ribityl-3,4-xylydino)uracil (21), prepared *in situ* from the condensation of 6-chloro-5-nitrouracil (22) with *N*-D-ribityl-3,4-xylydine (11), was hydrogenated over palladium on charcoal in acetic acid. The filtrate included 5-amino-6-(*N*-D-ribityl-3,4-xylydino)uracil (23) and was maintained at room temperature to precipitate (1) by autoxidation (35). These two pathways are suitable for the preparation of riboflavin analogues possessing several substituents (Fig. 4).

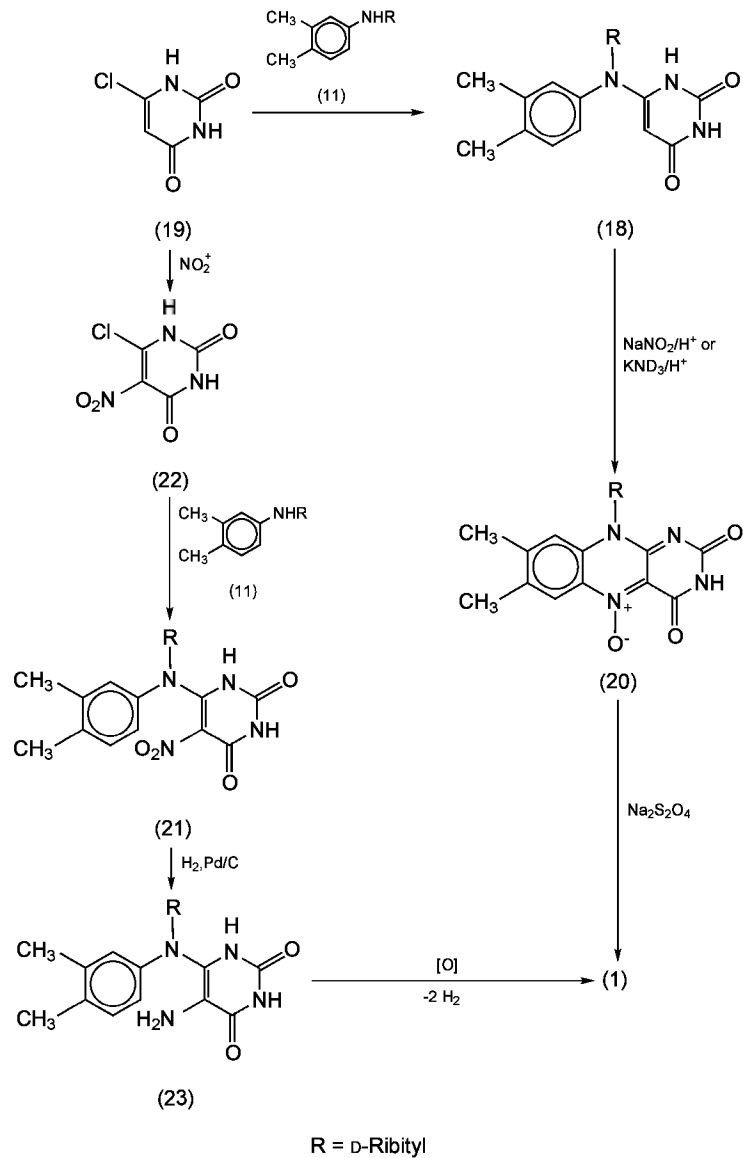


Fig. 4. Alternative syntheses of riboflavin.

The chemistry of flavins, including several synthetic methods for the preparation of *N*-D-ribityl-3,4-xylydine (11) is reviewed in Reference 36.

Microbial Synthesis

Biosynthetic Mechanism. Riboflavin is produced by many microorganisms, including *Ashbya gossypii*, *Asperigillus sp*, *Eremothecium ashbyii*, *Candida* yeasts, *Debaryomyces* yeasts, *Hansenula* yeasts, *Pichia* yeasts, *Agrobacter* sp, *Clostridium* sp, and *Bacillus* sp. These organisms have been used frequently in the elucidation of the biosynthetic pathway (37,38). The mechanism of riboflavin biosynthesis has formally been deduced from data derived from several experiments involving a variety of organisms (Fig. 5). Included are conversion of a purine such as guanosine triphosphate (GTP) to 6,7-dimethyl-8-D-ribityllumazine (16) (39), and the conversion of (16) to (1). This concept of the biochemical formation of riboflavin was verified *in vitro* under nonenzymatic conditions (40) (see MICROBIAL TRANSFORMATIONS).

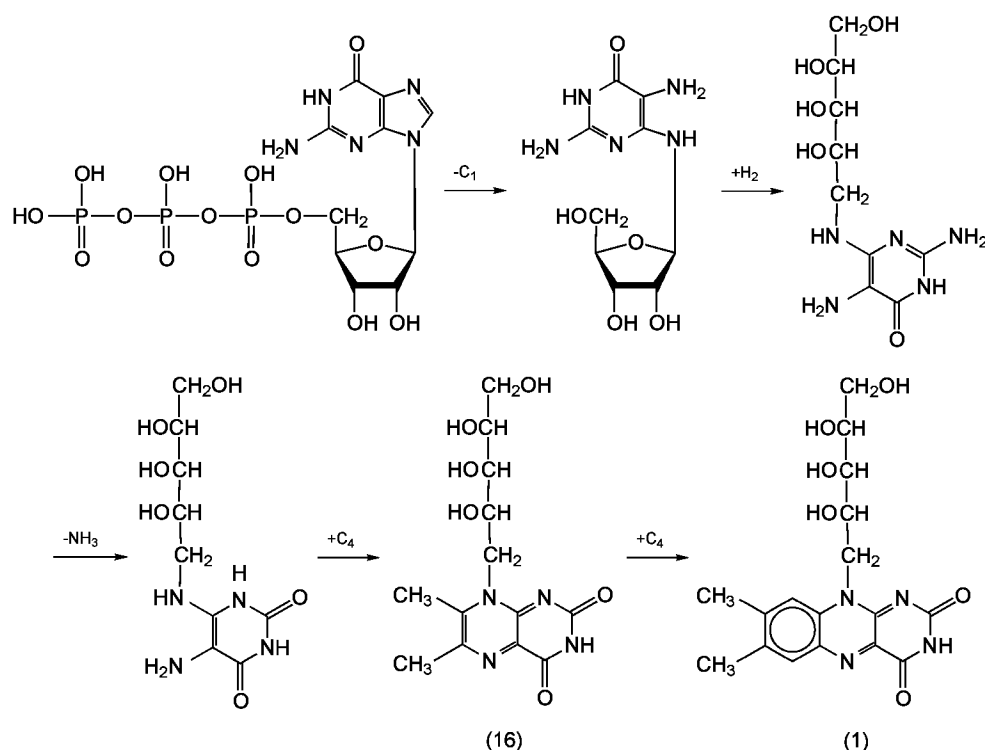


Fig. 5. Biosynthesis pathway to riboflavin.

Fermentative Manufacture. Throughout the years, riboflavin yields obtained by fermentation have been improved to the point of commercial feasibility. Most of the riboflavin thus produced is consumed in the form of crude concentrates for the enrichment of animal feeds. Riboflavin was first produced by fermentation in 1940 from the residue of butanol–acetone fermentation. Several methods were developed for large-scale production (41). A suitable carbohydrate-containing mash is prepared and sterilized, and the pH adjusted to 6–7. The mash is buffered with calcium carbonate, inoculated with *Clostridium acetobutylicum*, and incubated at 37–40°C for 2–3 d. The yield is ca 70 mg riboflavin L (42) (see FERMENTATION).

Most varieties of *Candida* yeasts produce substantial amounts of riboflavin when glucose is the carbon source. Particularly, *Candida guilliermondii* and *Candida flaveri* produce high yields on a simple synthetic medium of low cost. Some modifications employing several *Candida* yeasts have been patented; eg, *C. intermedia* var. *A*, a newly isolated microorganism assimilating lactose and ethanol, gave yields of 49.2 mg riboflavin L from ethanol in the presence of biotin (43). *Candida T-3*, assimilating methanol, produces riboflavin by this method from 50 g methanol (44). *Candida* bacteria have the advantage of extremely low iron tolerance. Chelating agents, such as 2,2'-dipyridyl, are recommended to control the iron content (45).

Most of the commercial riboflavin production by aerobic fermentation is obtained by biosynthesis with the yeastlike fungus *Eremothecium ashbyii*. Many variations for the production of riboflavin by *E. ashbyii* have been patented. Employing *E. ashbyii* grown on a yeast medium, riboflavin production on an industrial scale is said to have reached 11.6 g kg from dried powder (46). Riboflavin is also obtained by fermentation with *E. ashbyii* preserved on Difco-treated millet seeds. The culture, kept for 7–8 d at 32°C on the medium containing crude collagen, corn extracts, unrefined plant oil, glucose, KH₂PO₄, trace elements, and H₂O at pH 7–8, gave 4.5–5.0 g L (47). In another procedure, *E. ashbyii* was cultivated on a culture medium containing sources of assimilable nitrogen, essential minerals, and growth factors, along with sources of assimilable carbohydrates, unsaturated fatty acids, saccharides, and amino acid or their salts. Incubation at 29°C for 6 d on a rotary shaker with aeration gave average yields of 3.8 g L (48). *Eremothecium ashbyii* grown in a culture medium containing the oil cake obtained after extraction of lipids from the biomass grown on hydrocarbons gives an even better yield (49).

In operations similar to the *E. ashbyii* procedures, the closely related fungus *Ashbya gossypii* gave similar yields. Thus, a yield of 7.3 g L was obtained with a lyophilized culture in a medium containing fat, leather glue, and corn extracts (50), and 6.420 g L with bone or hide fat, alone or in a mixture with other plant or animal fats as the carbon source (51). The yield from immersed cultures of *A. gossypii* was increased to 6.93–7.20 g L by use of waste fats or technical cod-liver oil (52).

Riboflavin is also made by aerobic culturing of *Pichia guilliermondii* on a medium containing *n*-C₁₀–C₁₅ paraffins in a yield of 280.5 mg L (53). A process employing *Pichia* yeasts, such as *P. miso*, *P. miso Mogi*, or *P. mogi*, in a medium containing a hydrocarbon as the carbon source, has been patented (54).

Processes employing *Torulopsis xylinus* (55), *Hansenula polymorpha* (56), *Brevibacterium ammoniagenes* (57), *Achromobacter butryi* (58), *Micrococcus lactis* (59), *Streptomyces testaceus* (60), and others have also been patented. These procedures yield, at most, several hundred milligrams of riboflavin per liter.

Manufacturing procedures of riboflavin have also appeared using *Saccharomyces* bacteria, eg, fermentation with a purine-independent *S. reverse* mutant (61) and with *S. cerevisiae* NH-268 (62) produced 2.79 g L and 4.9 g L, respectively.

Further efficient fermentative methods for manufacture of riboflavin have been patented; one is culturing *C. famata* by restricting the carbon source uptake rate, thereby restricting growth in a linear manner by restriction of a micronutrient. By this method, productivity was increased to >0.17 g riboflavin L h (63). The other method, using *Bacillus subtilis* AJ 12644 low in guanosine monophosphate hydrolase activity, yielded crude riboflavin 0.9 g L 3 days, when cultured in a medium including soy protein, salts, and amino acids (64).

In recent (ca 1997) years, fermentative manufacturing methods using recombinant microorganisms have been developed. The mutant clone GA18Y8-6#2, prepared by fusion and mutagenation of *C. flaveri* mutants A22 and GA18, produced riboflavin 7.0–7.5 g L 6 d, when cultivated in 4B medium with a supplement of FeCl₃, yeast extracts, peptone, and malt extracts (65). Riboflavin overproducing bacteria prepared by expression of the cloned *rib* operon of *Bacillus subtilis* showed increases in riboflavin manufacture of up to a hundredfold (up to 0.7 g riboflavin L 48 h). The *rib* operon was cloned as series of overlapping genes using an oligonucleotide derived from the amino acid sequence of the riboflavin synthase β subunit (66). Culturing recombinant *Corynebacterium ammoniagenes* KY13313 harboring gene for at least *guanidine triphosphate cyclohydrolase* and riboflavin synthase produced riboflavin ~30-fold higher than that with the controlled bacteria (67).

Industrial Aspects

For the industrial production of riboflavin as pharmaceuticals, the traditional methodology comprising the direct condensation of (13) with (14) in an acidic medium with continuous optimization of the reaction conditions is still used (28). A great part of riboflavin manufactured by fermentative methods is used for feeds in the form of concentrates. The present world demand of riboflavin may be about 2500 t per year. Of this amount, 60% o, 25% o, and 15% o are used for feeds, pharmaceuticals, and foodstuffs, respectively. The main producers are Hoffmann-La Roche, BASF, Merck & Co., and others.

Analytical Methods

Riboflavin can be assayed by chemical, enzymatic, and microbiological methods. The most commonly used chemical method is fluorometry, which involves the measurement of intense yellow-green fluorescence with a maximum at 565 nm in neutral aqueous solutions. The fluorometric determinations of flavins can be carried out by measuring the intensity of either the natural fluorescence of flavins or the fluorescence of lumiflavin formed by the irradiation of flavin in alkaline solution (68). The later development of a laser–fluorescence technique has extended the limits of detection for riboflavin by two orders of magnitude (69,70).

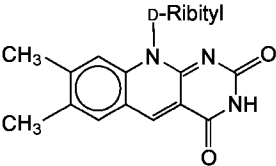
Polarography is applied in the presence of other vitamins, eg, in multivitamin tablets, without separation. The polarography of flavins is reviewed in Reference 71.

The microbial assay is based on the growth of *Lactobacillus casei* in the natural (72) or modified form. The lactic acid formed is titrated or, preferably, the turbidity measured photometrically. In a more sensitive assay, *Leuconostoc mesenteroides* is employed as the assay organism (73). It is 50 times more sensitive than *L. casei* for assaying riboflavin and its analogues (0.1 ng mL vs 20 ng mL for *L. casei*). A very useful method for measuring total riboflavin in body fluids and tissues is based on the riboflavin requirement of the protozoan ciliate *Tetrahymena pyriformis*, which is sensitive and specific for riboflavin. This method can be applied to large-scale nutrition studies.

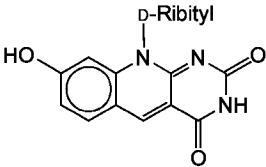
Although riboflavin can be assayed more readily by chemical or microbiological methods than by animal methods, the latter are preferred for nutritional studies and as the basis of other techniques. Such assays depend upon a growth response; the rat or chick is the preferred experimental animal. This method is particularly useful for assaying riboflavin derivatives, since the substituents frequently reduce or eliminate the biological activity.

An enzymatic method for assessing riboflavin deficiency in humans has been developed (74). It is based on the fact that NADPH-dependent glutathione reductase of red cells reflects riboflavin fluctuations.

High pressure liquid chromatography (hplc) has been extensively used for the riboflavin determinations. This method is usually automated and more rapid and sensitive than the microbial method. It has been used in combination with fluorometric detection for the riboflavin assay in foods (75), meat and meat products (76), and enriched and fortified foods (77), as well as in a simple assay for animal tissues (78). A rapid and efficient reversed-phase hplc method is described for the quantitative separation of flavin coenzymes and their structural analogues such as 5-deazaflavin [19342-73-5] (24) and 8-hydroxy-7-demethyl-5-deazariboflavin [37333-48-5] (25) (F₄₂₀ chromophore) (79). Comprehensive reviews of the analytical methods for riboflavin are given in References 80 and 81.



(24)



(25)

Biological Function

In biological systems, riboflavin functions almost exclusively in the form of flavoproteins, in which the FMN or FAD is generally bound as prosthetic group or coenzyme to specific proteins. These enzymes catalyze oxidation–reduction reactions (Table 2). The flavin group of the oxidized coenzyme is reduced chemically or enzymatically to 1,5-dihydroflavin coenzyme, probably in two one-electron steps, each involving the addition of a single electron. Stable semiquinone radicals are formed as intermediates, because the unpaired electron is highly delocalized by the conjugated isalloxazine structure.

Table 2. Some Reactions Catalyzed by Flavoproteins^a

Enzyme	Electron donor	Product	Coenzyme and other components	Electron acceptor
D-amino acid oxidase	D-amino acids	α-keto acids +NH ₃	2FAD	O ₂ → H ₂ O ₂
L-amino acid oxidase (liver)	L-amino acids	α-keto acids +NH ₃	2FAD	O ₂ → H ₂ O ₂
L-amino acid oxidase (kidney)	L-amino acids	α-keto acids +NH ₃	2FMN	O ₂ → H ₂ O ₂
L(+)-lactate dehydro-genase (yeast)	lactate	pyruvate	1FMN; 1theme (<i>yt b₅</i>)	respiratory chain
glycolate oxidase	glycolate	glyoxylate	FMN	O ₂ → H ₂ O ₂
NADH-cytochrome <i>c</i> reductase	NADH	NAD ⁺	FMN; 2Mo, NHI	cytochrome c _{ox}
NADH-cytochrome <i>b₅</i> reductase	NADH	NAD ⁺	FAD; Fe	respira-tory chain
aldehyde oxidase (liver)	aldehydes	carboxylic acids	FAD; Fe, Mo	cytochrome <i>b₅</i>
α-glycerol phosphate dehydrogenase	glycerol 3-phos-phate	dihydroxy-acetone phosphate	FAD; Fe	respiratory chain
succinate dehydro-genase	succinate	fumarate	FAD; Fe, NHI	respiratory chain
acyl-CoA(C –C ₁₂) dehydrogenase	acyl-CoA	enoyl-CoA	FAD	electron-transferring flavoprotein
nitrate reductase	NADPH	NADP ⁺	FAD; Mo, Fe	nitrate
nitrite reductase	NADPH	NADP ⁺	FAD; Mo, Fe	nitrite
xanthine oxidase	xanthine	uric acid	FAD; Mo, Fe	O ₂
lipoate dehydrogenase	reduced lipoic acid	oxidized lipoic acid	2FAD	NAD ⁺
dehydroorotate dehydrogenase	dihydroorotic acid	orotic acid	2FMN; 2FAD, 4Fe	undetermined

^a Ref. 82.

In contrast to the nicotinamide nucleotide dehydrogenases, the prosthetic groups FMN and FAD are firmly associated with the proteins, and the flavin groups are usually only separated from the apoenzyme (protein) by acid treatment in water. However, in several covalently bound flavoproteins, the enzyme and flavin coenzymes are covalently affixed. In these cases, the flavin groups are isolated after the proteolytic digestion of the flavoproteins.

Many flavoproteins react directly with molecular oxygen to produce hydrogen peroxide. Some flavoproteins, such as the flavin-containing monooxygenase, give water instead of hydrogen peroxide. In these cases, one atom of oxygen is introduced into a substrate to undergo hydroxylation, whereas the other oxygen atom is released as water. Several flavoproteins include metal complexes where these reactions take place. A number of reviews of the preparation, properties, and mechanism of action of these enzymes have been published (83,84).

Deficiency, Requirements, and Toxicity

Riboflavin is essential for mammalian cells. A lack of riboflavin in the human diet causes characterized deficiency syndromes, such as sore throat, hyperemia, cheilosis, angular stomatitis, glossitis (magenta tongue), a generalized seborrheic dermatitis, scrotal and vulval skin changes, and a normocytic anemia. Because riboflavin is essential to the functioning of vitamins B and niacin, some symptons attributed to riboflavin deficiency are actually due to the failure of systems requiring these other nutrients to operate effectively (85).

The 1989 Recommended Dietary allowances (RDA) of the Food and Nutrition Board (86) are 0.6 mg riboflavin per 239 kJ (1000 kcal) for essentially healthy people of all ages. This leads to the ranging from 0.4 mg day for early infants to 1.8 mg day for young males. For elderly people and others whose daily calorie intake may be less than 478 kJ (2000 kcal), a minimum of 1.2 mg day is recommended. During pregnancy, an additional riboflavin intake of 0.3 mg day is recommended in view of the increased tissue synthesis for both fetal and maternal development. For the lactating woman, the requirement is assumed to increase by an amount at least equal to that excreted in milk, which has a mean riboflavin content of 35 µg 100 mL. At an average milk production of 750 mL day and 600 mL d during the first and second 6 months of lactation, riboflavin secretion is 0.26 mg d and 0.21 mg d, respectively. Because the utilization of the additional riboflavin for milk production is assumed to be 70%, and the coefficient of variation of milk production is 12.5%, an additional daily intake of 0.5 mg is recommended for the first 6 mo of lactation and 0.4 mg thereafter.

Riboflavin is essentially nontoxic. The LD₅₀ values in mice and rats by intraperitoneal injection are 340 mg kg (87) and 560 mg kg (88), respectively. The oral administration of 10 g kg to rats or 2 g kg to dogs showed no toxic effects (89).

Derivatives

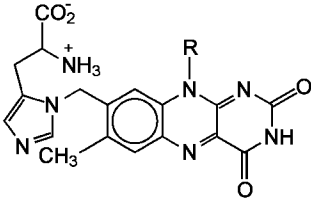
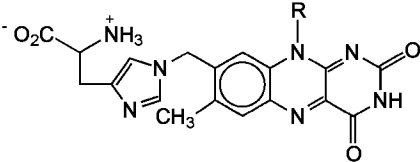
Riboflavin-5'-Phosphate. Riboflavin-5'-phosphate [146-17-8] (vitamin B₂ phosphate, flavin mononucleotide, FMN, cytoflav), C₁₇H₂₁N₄O₉P, mol wt 456.35, is a microcrystalline yellow solid, mp 195°C, [α]_D²⁸ +44.5° (2% soln in conc HCl) with biological and enzymatic activity. It is prepared by phosphorylation of riboflavin with chlorophosphoric acid (90), pyrophosphoric acid (91), metaphosphoric acid (92), or catechol cyclic phosphate (93). It is soluble in water to the extent of 3 g 100 mL at 25°C as the sodium salt but tends to gel. Because of the high sensitivity of FMN to uv, it must be preserved in dark, tight containers.

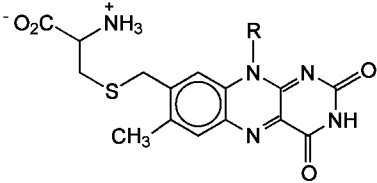
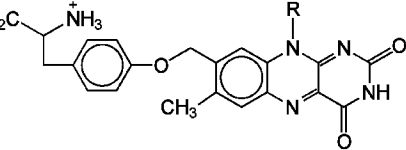
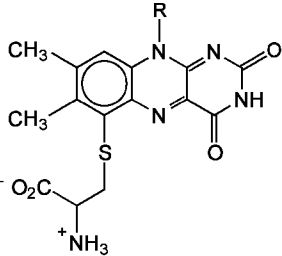
Flavin mononucleotide was first isolated from the yellow enzyme in yeast by Warburg and Christian in 1932 (4). The yellow enzyme was split into the protein and the yellow prosthetic group (coenzyme) by dialysis under acidic conditions. Flavin mononucleotide was isolated as its crystalline calcium salt and shown to be riboflavin-5'-phosphate; its structure was confirmed by chemical synthesis by Kuhn and Rudy (94). It is commercially available as the monosodium salt dihydrate [6184-17-4], with a water solubility of more than 200 times that of riboflavin. It has wide application in multivitamin and B-complex solutions, where it does not require the solubilizers needed for riboflavin.

Riboflavin-5'-Adenosine Diphosphate. Riboflavin-5'-adenosine diphosphate [146-14-5] (flavin-adenine dinucleotide, FAD), C₂₇H₃₃N₉O₁₅P₂ (2), mol wt 785.56, was first isolated in 1938 from the D-amino acid oxidase as its prosthetic group (95), where it was postulated to be flavin-adenine dinucleotide. The structure was established by the first synthesis in Todd's laboratory (96); the monosilver salt of FMN was condensed with 2',3'-isopropylidene-adenosine-5'-benzylphosphorchloridate, followed by removal of protective groups. It was also synthesized directly from FMN and adenosine-5'-monophosphate (AMP) with di-*p*-tolylcarbodiimide as the condensation agent (97). Another direct synthesis was achieved by dehydration between FMN and AMP with trifluoroacetic acid anhydride (98). A 40% yield was obtained by condensation of adenosine-5'-phosphoramidate and FMN using a mixture of pyridine and *o*-chlorophenol as the solvent (99). Condensation of AMP with FMN in ethoxyacetylene gives a 10–15% yield; by-products such as riboflavin-4',5'-cyclic phosphate are avoided (100). An efficient procedure was developed, which transforms the free acid of both nucleotides into their tri-*n*-butyl-ammonium salts and uses a reactive carbodiimide for the coupling reaction. The yield of FAD can be as high as 70% (101). In addition to D-amino acid oxidase, FAD is the prosthetic group for the other flavoproteins including glucose oxidase, lycine oxidase, fumarate hydrogenase, histaminase, and xanthine oxidase.

Covalently Bound Flavins. The FAD prosthetic group in mammalian succinate dehydrogenase was found to be covalently affixed to protein at the 8 α-position through the linkage of 3-position of histidine (102,103). Since then, several covalently bound riboflavins (104,105) have been found successively from the enzymes listed in Table 3. The biosynthetic mechanism, however, has not been clarified.

Table 3. Covalently Bound Flavins

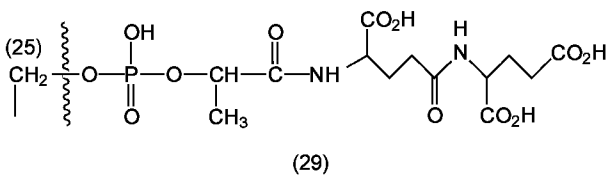
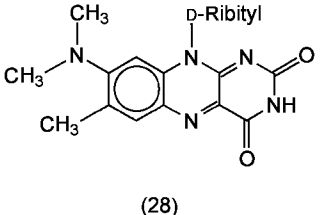
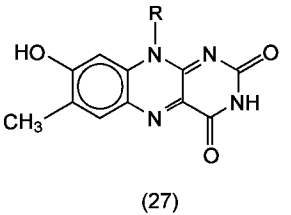
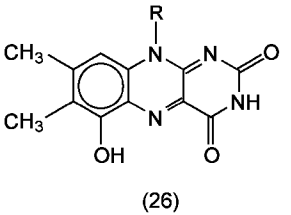
Prosthetic group (R D-Ribityl)	CAS Registry Number	Formula	Enzyme
	[58525-92-1]	C ₂₃ H ₂₇ C ₇ O ₈	thiamine dehydrogenaseβ-cyclopiazonate oxidocyclaseL-gulono-γ-lactone oxidaseL-galactonolactone oxidasecholesterol oxidase
8 α-(N ¹ -histidyl)riboflavin	[37854-44-7]	C ₂₃ H ₂₇ N ₇ O ₈	succinate dehydrogenasefumarate reductasesalcosine dehydrogenasesalcosine oxidasedimethylglycine dehydrogenaseD-hydroxynicotine oxidasecholine oxidaseD-gluconate dehydrogenase
	[35836-22-7]	C ₂₀ H ₂₅ N ₅ O ₈ S	monoamine oxidase Amonoamine
8 α-(N ³ -histidyl)riboflavin			

			oxidase B <i>Chlorobium</i> cytochrome C ₅₅₂ <i>Chlomatium</i> cytochrome C ₅₅₃
8 α-S-cysteinylriboflavin	[38065-74-6]	C ₂ H ₂₉ N ₅ O ₉	p-cresol methylhydroxylase
			
8 α-O-tyrosylriboflavin	[73647-60-6]	C ₂₀ H ₂₅ N ₅ O ₈ S	trimethylaminedehydrogenasedimethyl aminodehydrogenase
			
6-S-cysteinylriboflavin			

6-Hydroxyriboflavin. This compound [86120-61-8] (26) was isolated as a green coenzyme of the NADH dehydrogenase from *Peptostreptococcus elsdenii* and also from glycolate oxidase of porcine liver. It is not fluorescent, and its structure was established by synthesis (106). The 5'-monophosphate serves as a cofactor for glycolate oxidase from pig liver.

8-Nor-8-hydroxyriboflavin. A prosthetic group of red color has been isolated from NADH dehydrogenase of the electron-transferring flavoprotein of *Peptostreptococcus elsdenii*. Its structure [52134-62-0] (27) has been established as the FAD derivative of 8-hydroxy-7-methylisoalloxazine. Proof has been obtained by the synthesis of 8-hydroxy-7-methylisoalloxazine models and stepwise degradation of the naturally occurring compound (107).

Roseflavin. Roseflavin [51093-55-1], C₁₈H₂₃N₅O , (28), mol wt 405.41, mp 274–297°C, [α]_D = – 320° (0.1M NaOH), was isolated from a culture medium of *Streptomyces davawensis* as dark, reddish-brown fine needles (from ethanol); the 8-methyl group of riboflavin is substituted by a dimethylamino group. This structure was confirmed by the synthesis. Roseflavin shows antimicrobial activity against gram-positive bacteria (108).



5-Deazariboflavin. In 5-deazariboflavin (24), the N-5 of riboflavin is replaced by CH; it serves as cofactor for several flavin-catalyzed reactions (109). It was first synthesized in 1970 (110); improved synthetic processes were reported later (111).

A low potential electron carrier, the fluorescent factor F₄₂₀ [37333-48-5, 64885-97-8] (29) (it absorbs maximally at 420 nm), possessing a

5-deazaflavin moiety, was isolated from methane-producing bacteria (112). F₄₂₀ is an obligate intermediate for passage of an electron from H₂ to NADP⁺ to generate NADPH. The structure of F₄₂₀ was proposed as an 8-hydroxy-7-demethyl-5-deazaflavin derivative (25) (113) and confirmed as compound (29) by the total synthesis (114).

BIBLIOGRAPHY

"Riboflavin" in *ECT* 1st ed., Vol. 11, pp. 749–759, by G. E. Sita, Merck & Co.; "Riboflavine" in *ECT* 2nd ed., Vol. 17, pp. 445–458, by E. De Ritter, Hoffmann-La Roche, Inc.; "Riboflavin (B₂)" under "Vitamins" in *ECT* 3rd ed., Vol. 24, pp. 108–124, by F. Yoneda, Kumamoto University.

1. R. Kuhn, P. Gyntgy, and T. Wagner-Jauregg, *Ber. Dtsch. Chem. Ges.* **66**, 576 (1933).
2. *Idem. ibid.*, **66**, 317 (1933).
3. P. Karrer and K. Schnupp, *Helv. Chem. Acta* **17**, 735, 771 (1934).
4. O. Warburg and W. Christian, *Biochem. Z.* **254**, 438 (1932); *idem. ibid.* **266**, 377 (1933).
5. R. Kuhn, H. Ruby, and T. Wagner-Jauregg, *Ber. Dtsch. Chem. Ges.* **66**, 1950 (1933).
6. P. Karrer, K. Schnupp, and F. Benz, *Helv. Chim. Acta* **18**, 426, 522 (1935).
7. R. Kuhn, K. Reinemund, H. Kaltschmitt, K. Stnbele, and H. Trischmann, *Naturwissenschaften* **23**, 260 (1935).
8. E. A. Woodcock and J. J. Warthesen, *J. Food Sci.* **47**, 545 (Apr. 1982).
9. **U.S. Pat. 2,407,412** (Sept. 10, 1946), D. V. Frost (to Abbott Laboratories).
10. A. W. Varnes, R. B. Dodson, and E. L. Wehry, *J. Am. Chem. Soc.* **94**, 946 (1972).
11. P. Karrer, H. Salmon, K. Schnupp, E. Schlittler, and H. Fritsche, *Helv. Chim. Acta* **17**, 1010 (1934).
12. E. C. Smith and D. Metzler, *J. Am. Chem. Soc.* **85**, 3285 (1963).
13. R. Kuhn and T. Wagner-Jauregg, *Ber. Dtsch. Chem. Ges.* **67**, 361 (1934).
14. L. Michaelis, M. P. Schubert, and C. V. Smythe, *J. Biol. Chem.* **116**, 587 (1936).
15. A. Ehrenberg, *Acta Chem. Scand.* **11**, 205 (1957).
16. B. Commoner and B. Lippincott, *Proc. Natl. Acad. Sci. U.S.A.* **44**, 1110 (1958).
17. F. Møller, P. Hemmerich, and A. Ehrenberg in H. Kamin, ed., *Flavins and Flavoproteins*, University Park Press, Baltimore, Md., 1971, p. 107.
18. F. Møller and co-workers, *Eur. J. Biochem.* **116**, 17 (1981).
19. H. Kurreck and co-workers, *J. Am. Chem. Soc.* **106**, 737 (1984).
20. P. Hemmerich, *Experientia* **16**, 534 (1969).
21. P. Bamberg and P. Hemmerich, *Helv. Chim. Acta* **44**, 1001 (1961); K. H. Dubley, A. Ehrenberg, P. Hemmerich, and F. Møller, *Helv. Chim. Acta* **47**, 1354 (1964).
22. A. Ehrenberg and P. Hemmerich, in T. P. Singer, ed., *Biological Oxidations*, Wiley-Interscience, New York, 1968, p. 722.
23. R. Kuhn and F. Weygand, *Ber. Dtsch. Chem. Ges.* **68**, 1282 (1935).
24. P. Karrer and H. F. Meerwein, *Helv. Chim. Acta* **18**, 1130 (1935).
25. *Idem. ibid.*, **19**, 264 (1936).
26. M. Tishler, J. W. Wellman, and K. Ladenburg, *J. Am. Chem. Soc.* **67**, 2165 (1945).
27. T. Matsukawa and K. Shirakawa, *Yakugaku Zasshi* **69**, 208 (1949).
28. M. Tishler, K. Pfister, R. D. Babson, K. Ladenburg, and A. J. Fleming, *J. Am. Chem. Soc.* **69**, 1487 (1947).
29. **U.S. Pat. 4,673,742** (June 16, 1987), J. Grimmer and H. C. Horn (to BASF AF); **Ger. Pat. 3,542,837** (June 11, 1987), J. Grimmer (to BASF AG).
30. F. Bergel, A. Cohen, and J. W. Haworth, *J. Chem. Soc.*, 165 (1945).
31. R. M. Cresswell and H. C. S. Wood, *J. Chem. Soc.*, 4768 (1960).
32. **Jpn. Pat. 10,031** (Nov. 13, 1959), A. Masuda (to Takeda Pharmaceutical Industries).
33. J. Davol and D. D. Evans, *J. Chem. Soc.*, 5041 (1960); see also R. M. Cresswell, T. Neilson, and H. C. S. Wood, *J. Chem. Soc.*, 477 (1961).
34. F. Yoneda, Y. Sakuma, M. Ichiba, and K. Shinomura, *J. Am. Chem. Soc.* **98**, 830 (1976).
35. F. Yoneda, Y. Sakuma, and K. Shinozuka, *J. Chem. Soc. Perkin Trans. 1*, 348 (1978).
36. T. Wagner-Jauregg in W. H. Sebrell, Jr., and R. E. Harris, eds., *The Vitamins*, Vol. V, Academic Press, Inc., New York, 1972, p. 19; J. P. Lambooy in R. C. Elderfield, ed., *Heterocyclic Compounds*, Vol. 9, John Wiley & Sons, Inc., New York, 1967, p. 118.
37. G. W. E. Plaut in M. Florkin and E. H. Stotz, eds., *Comprehensive Biochemistry*, Vol. 21, Elsevier Publishing Co., Amsterdam and New York, 1971, p. 11.
38. A. Bacher in F. Møller, ed., *Chemistry and Biochemistry of Flavoenzymes*, Vol. 1, CRC Press, Boca Raton, Fla., 1991, p. 215.
39. M. Mitsuda, K. Nakajima, and T. Nadamoto, *J. Nutr. Sci. Vitaminol.* **22**, 477 (1976); **23**, 71 (1977).
40. T. Rowan and H. C. S. Wood, *J. Chem. Soc.*, 452 (1968); T. Paterson and H. C. S. Wood, *J. Chem. Soc. Perkin Trans. 1*, 1051 (1972).
41. **U.S. Pat. 2,202,161** (May 28, 1940), C. S. Miner (to Commercial Solvents Corp.).
42. **U.S. Pat. 2,369,680** (Feb. 20, 1945), R. E. Maeda, H. L. Polland, and N. E. Rodgers (to Western Condensing Co.); **U.S. Pat. 2,449,144** (Sept. 14, 1948), N. E. Rogers, H. L. Polland, and R. E. Maeda (to Western Condensing Co.).
43. **Jpn. Pat. 73 19,958** (June 18, 1973), S. Sugawara and K. Sato (to Nippon Beet Sugar Manufacturing Co.).
44. **Jpn. Pat. 76 19,187** (Feb. 16, 1976), Y. Ichida, H. Abe, and A. Aoike (to Kuraray Co.).
45. **U.S. Pat. 2,425,280** (Aug. 5, 1947), R. Hickey (to Commercial Solvents Corp.).
46. **Ger. Pat. 1,936,238** (Jan. 28, 1971), E. M. Dikanskaya and A. A. Balabanova (to All-Union Scientific Research Institute of Protein Biosynthesis).
47. **Ger. Pat. 2,028,355** (Jan. 14, 1971), I. Nitelea and co-workers (to Romania Antibiotic Plant).
48. **Ger. Pat. 1,767,260** (Aug. 19, 1976), G. M. Miescher (to Commercial Solvents Corp.).
49. USSR Pat. 194,261 (June 15, 1974), E. M. Dikanskaya (to All-Union Scientific-Research Institute of Protein Biosynthesis).
50. **Ger. Pat. 2,453,827** (May 15, 1975), T. Slave and co-workers (to Institutul de Cercetari Chimico Farmaceutice).
51. Pol. Pat. 66,611 (Mar. 15, 1973), T. Szczesniak and co-workers (to Instytut Przemyslu Fermentacyjnego).
52. Pol. Pat. 76,481 (Mar. 10, 1975), T. Szeszaniak and co-workers (to Instytut Przemyslu Farmaceutycznego).
53. **Ger. Pat. 2,037,905** (Feb. 3, 1972), T. Kamikubo and N. Hiroshima (to Kanegafuchi Chemical Industry Co.).
54. **U.S. Pat. 3,433,707** (Mar. 18, 1969), T. Matsubayashi and Y. Suzuki (to Dai Nippon Sugar Manufacturing Co. and Nitto Physico-Chemical Research Institute).
55. **Jpn. Pats. 73 96,790 and 73 96 791** (Dec. 10, 1973), T. Fukukawa, T. Matsuyoshi, and J. Hiratsuka (to Mitsui Petrochemical Industries).
56. **Jpn. Pat. 79 80,495** (June 27, 1979), S. Uragami (to Mitsubishi Gas Chemical Co.).
57. **Jpn. Pat. 77 110,897** (Sept. 17, 1977), K. Nakayama, K. Araki, and S. Shimojo (to Kyowa Hakko Kogyo Co.).
58. **Jpn. Pat. 77 54,094** (May 2, 1977), I. Chibata and co-workers (to Tanabe Seiyaku Co.).
59. USSR Pat. 511,742 (Sept. 25, 1976), T. E. Popova and co-workers (to All-Union Scientific-Research Institute of Protein Biosynthesis).
60. **Jpn. Pat. 75 116,690** (Sept. 12, 1975), H. Umezawa and co-workers (to Sanraku-Ocean Co.).
61. **Jpn. Kokai Tokkyo Koho Rus. Pat. 87 25,996** (Feb. 3, 1987), A. Matsuyama and co-workers (to Daicel Chemical Industries, Ltd.).
62. **Jpn. Kokai Tokkyo Koho Rus. Pat. 88 112,996** (May 18, 1988), A. Matsuyama and co-workers (to Daicel Chemical Industries, Ltd.).
63. **PCA Int. Appl. WO 9201,060** (Jan. 23, 1992), R. B. Bailey, G. W. Lauderdale, D. L. Heefner, C. A. Weaver, M. J. Yarus, L. A. Burdzinski, and A.

Boyte (to Coors Biotechnology, Inc.).

64. **Eur. Pat. Appl. EP 531,708** (Mar. 17, 1993), N. Usui, Y. Yamamoto, and T. Nakamatu (to Ajinomoto Co.).

65. **Eur. Pat. Appl. EP 231,605** (Aug. 12, 1987), D. L. Heefner, M. Yarus, A. Boyts, and L. Burdzinski (to Adolph Coors Co.).

66. **Eur. Pat. Appl. EP 405,370** (Jan. 2, 1991), D. B. Perkins, J. K. Pero, and A. Sloma (to Hoffmann-La Roche, F., Co.).

67. **Eur. Pat. Appl. EP 604,060** (Jan. 29, 1994), S. Koizumi, Y. Yonetani, and S. Teshiba (to Kyowa Hakko Kogyo Co.).

68. J. Koziol in D. B. McCormick and L. D. Wright, eds., *Methods in Enzymology*, Vol. 43, Academic Press, Inc., New York, 1971, p. 253.

69. J. H. Richardson in D. B. McCormick and L. D. Wright, eds., *Methods in Enzymology*, Vol. 66, Academic Press, Inc., New York, 1980, p. 416.

70. N. Ishibashi, T. Ogawa, T. Imasaka, and M. Kunitake, *Anal. Chem.* **51**, 2096 (1979).

71. E. Knobloch in Ref. 63, p. 305.

72. E. E. Snell and F. M. Strong, *Ind. Eng. Chem. Anal. Ed.* **11**, 346 (1939).

73. H. A. Kornberg, R. S. Langdon, and V. H. Cheldelin, *Anal. Chem.* **20**, 81 (1948).

74. J. A. Tillotson and E. M. Baker, *Am. J. Clin. Nutr.* **25**, 425 (1972).

75. P. J. Richardson, D. J. Favell, G. C. Gidley, and A. D. Jones, *Proc. Anal. Div. Chem. Soc.* **15**, 53 (1978).

76. C. Y. Wang and F. A. Moseley, *J. Agric. Food Chem.* **28**, 483 (1980).

77. J. F. Kamman, T. P. Labuza, and J. J. Warthesen, *J. Food Sci.* **45**, 1497 (1980).

78. K. Yagi and M. Sato, *Biochem. Int.* **2**, 327 (1981).

79. D. R. Light, C. Walsh, and M. A. Marletta, *Anal. Biochem.* **109**, 87 (1980).

80. W. N. Pearson in P. Gyurgy and W. N. Pearson, eds., *The Vitamins*, Vol. VII, Academic Press, Inc., New York, 1967, p. 99.

81. H. Baker and O. Frank in R. S. Rivlin, ed., *Riboflavin*, Plenum Press, New York, 1975, p. 49.

82. E. E. Conn, P. K. Stumpf, G. Bruening, and R. H. Doi, *Outlines of Biochemistry*, 5th ed., John Wiley & Sons, Inc., New York, 1987, p. 207.

83. K. M. Horowitz and L. A. Wittig in W. H. Sebrell, Jr. and R. E. Harris, eds., *The Vitamins*, Vol. V, Academic Press, Inc., New York, 1972, p. 53; C. A. Hamilton, *Prog. Bioorg. Chem.* **1**, 83 (1971); T. C. Bruice, *Prog. Bioorg. Chem.* **4**, 1 (1976); C. Walsh, *Annu. Rev. Biochem.* **47**, 881 (1978); V. Massey and P. Hemmerich in P. D. Boyer, ed., *The Enzymes*, Vol. 12, Academic Press, Inc., New York, 1976, p. 191.

84. F. Møller, ed., *Chemistry and Biochemistry of Flavoenzymes*, Vol. 1, 1991; Vol. 2, 1991; and Vol. 3, 1992 CRC Press, Boca Raton.

85. D. B. McCormick in M. E. Sils and V. R. Young, eds., *Modern Nutrition in Health and Disease*, Lea and Febiger, Philadelphia, Pa., 1988, p. 362.

86. *Recommended Dietary Allowances*, 10th ed., Food and Nutrition Board, National Research Council, National Academy Press, Washington, D.C., 1989, p. 132.

87. R. Kuhn and P. Boulanger, *Hoppe-Seyler's Z. Physiol. Chem.* **241**, 233 (1936).

88. K. Unna and J. G. Greslin, *J. Pharmacol. Exp. Ther.* **76**, 75 (1942).

89. V. Demole, *Z. Vitaminforsch.* **7**, 138 (1938).

90. **U.S. Pats. 2,610,178 and 2,610,179** (Sept. 9, 1952), L. A. Flexser and W. G. Farkas (to Hoffmann-La Roche Inc.).

91. **U.S. Pat. 2,535,385** (Dec. 26, 1950), P. J. Breivogel (to White Laboratories).

92. M. Viscontini, C. Ebnother, and P. Karrer, *Helv. Chim. Acta* **35**, 457 (1952); M. Viscontini and co-workers, *Helv. Chem. Acta* **38**, 15 (1955).

93. T. Ukita and K. Nagasawa, *Chem. Pharm. Bull.* **7**, 465 (1959).

94. R. Kuhn and H. Rudy, *Ber. Dtsch. Chem. Ges.* **68**, 353 (1935).

95. O. Warburg and W. Christian, *Biochem. Z.* **296**, 294 (1938); **298**, 150 (1938); O. Warburg, W. Christian, and A. Griese, *Biochem. Z.* **297**, 417 (1938).

96. S. M. H. Christie, G. W. Kenner, and A. R. Todd, *Nature (London)* **170**, 924 (1952); *J. Chem. Soc.*, 46 (1954).

97. F. M. Huenekens and G. L. Kilgour, *J. Am. Chem. Soc.* **77**, 6716 (1955).

98. C. DeLuca and N. O. Kaplan, *J. Biol. Chem.* **223**, 569 (1956).

99. J. G. Moffatt and H. G. Khorana, *J. Am. Chem. Soc.* **80**, 3756 (1958).

100. H. Wassermann and D. Cohen, *Chem. Eng. News*, 47 (1962).

101. F. Cramer and H. Neunhoffer, *Ber. Dtsch. Chem. Ges.* **95**, 1664 (1962).

102. E. B. Kearney and T. P. Singer, *Biochem. Biophys. Acta* **17**, 596 (1955).

103. T. Y. Wang, C. L. Tsuo, and Y. L. Wang, *Sci. Sin.* **5**, 73 (1956).

104. T. P. Singer and D. E. Edmondson in Ref. 64, p. 253 and references cited therein.

105. D. E. Edmondson and R. De Francesco in Ref. 37, p. 73.

106. S. G. Mayhew, C. D. Whitefield, S. Ghisla, and M. Schuman-Juems, *Eur. J. Biochem.* **44**, 579 (1974).

107. S. Ghisla and S. G. Mayhew, *Eur. J. Biochem.* **63**, 373 (1976).

108. S. Otani, M. Takatsu, M. Nakano, S. Kasai, R. Miura, and K. Matsui, *J. Antibiot.* **27**, 88 (1974); S. Kasai, R. Miura, and K. Matsui, *Bull. Chem. Soc. Jpn.* **48**, 2877 (1975).

109. P. Hemmerich, V. Massey, and H. Fenner, *FEBS Lett.* **84**, 5 (1977).

110. D. E. O'Brien, L. T. Weinstock, and C. C. Cheng, *J. Heterocycl. Chem.* **7**, 99 (1970).

111. F. Yoneda in Ref. 64, p. 267; F. Yoneda and B. Kokel in Ref. 37, p. 121.

112. D. Eirich, G. D. Vogels, and R. S. Wolfe, *Biochemistry* **17**, 4583 (1978).

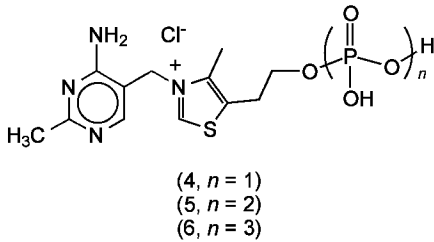
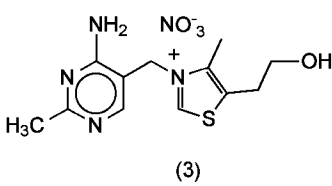
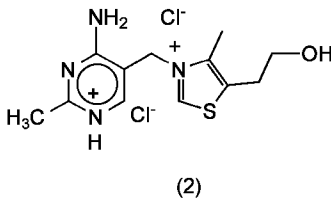
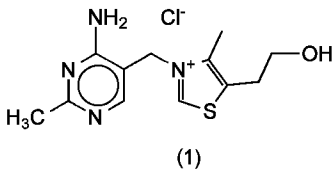
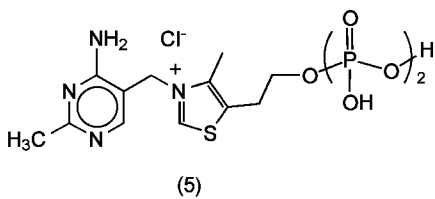
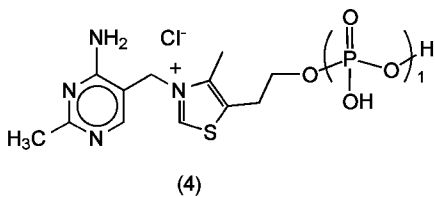
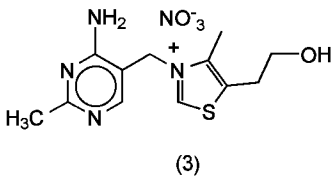
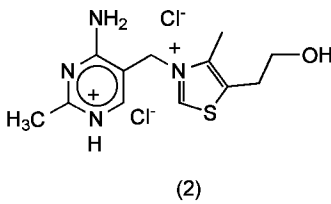
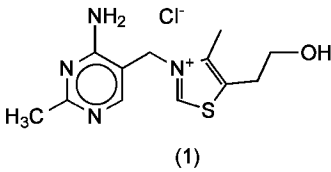
113. W. T. Ashton, R. D. Brown, F. Jacobson, and C. Walsh, *J. Am. Chem. Soc.* **101**, 4419 (1979).

114. T. Kimachi, M. Kawase, S. Matsuki, K. Tanaka, and F. Yoneda, *J. Chem. Soc., Perkin Trans. 1*, 253 (1990).

Fumio Yoneda
Fujimoto Pharmaceutical Corporation

THIAMINE (B₁)

Thiamine [59-43-8] is the official IUPAC-IUB name for 3-(4-amino-2-methyl-5pyrimidinyl) methyl-5-(2-hydroxyethyl)-4-methylthiazolium chloride, C₁₂H₁₇N₄OSCl (1). The names thiamine, thiamin (used in many official and commercial documents), aneurine, and the older term vitamin B₁ are also casually used. These names usually refer to the chloride hydrochloride [67-03-8] (2), the common biological and commercial form.



Thiamine is found in varying, low levels as its salts and phosphate esters in the tissues of practically all life forms, where its pyrophosphate [154-87-0] (5), known as cocarboxylase, plays essential roles in carbohydrate metabolism. Plants and most microorganisms biosynthesize thiamine, but animals are incapable of doing so, require it in their diets in amounts varying with their carbohydrate use, and excrete the excess. Thiamine content of foods varies and can be partially lost during storage or processing, as it is one of the less stable of the vitamins. Thiamine deficiencies can lead to a range of effects, from the malaise and fatigue of a lesser deficiency to the serious disorders of gross deficiency observed for many animal species, such as beriberi, where weight loss, heart disease, and serious neurological degeneration can lead to death. As preventive measures, thiamine is used to enrich foods and

feeds and as an ingredient of dietary supplements. Thiamine is also used therapeutically for treatment of specific deficiency conditions. Thiamine was one of the first of the vitamins to be manufactured for commerce. The common commercially available forms of thiamine, the chloride hydrochloride (2) and the mononitrate [532-43-4] (3), are manufactured by chemical processes operated at fine chemical scales. At present, natural sources and bioprocesses are not cost-competitive for bulk production.

The history of thiamine is linked with the disease beriberi, which once caused a great toll of suffering and death in parts of the world where milled or polished rice is a staple of the diet (1). Beriberi was recognized in China as early as 2600 BC, but the first cure was only demonstrated in 1885. Takaki, then surgeon general of the Japanese navy, eradicated beriberi in the fleet by adding more protein to the sailors' diet of mainly refined rice to achieve a more Western-style diet. By 1901 the Dutch physicians Eijkman and Grijn showed hens kept on a diet of de-husked rice developed a disease similar to human beriberi which could be reversed or prevented by adding rice bran and other nitrogenous materials to their feed. They suggested beriberi was caused not by pathogens or toxins but by the lack of a vitally important food constituent, which was a radical idea at the time. They also showed the preventive factor was leached by water and destroyed by heat. Searches for a specific water-soluble nitrogenous substance led the English chemist Funk in 1911 to coin the term *vitamine* (vital amine, an amine essential for life), popularizing the concept of deficiency diseases. In 1926, a small, pure crystalline sample of thiamine hydrochloride was painstakingly first isolated from rice bran extracts by Dutch chemists Jansen and Donath. After eight years of effort, in 1932 the German chemist Windaus and his co-workers obtained pure thiamine from yeast extracts and established the correct empirical formula. In 1935–1936, following similarly prolonged isolation efforts, the American chemists Williams and Cline and German chemist Grewe independently proposed the correct structure based on degradation studies. The name thiamine was first used by Williams. Shortly thereafter, the Williams group confirmed the structure by rational synthesis (2), followed closely by two different syntheses by other workers (3–5). Whereas in 1933 Williams had succeeded in isolating only gram quantities from metric tons of rice polishings, a highly enriched source, in 1937 chemists at Merck and Hoffmann-La Roche developed a production level of about 100 kg within a year based on two different, relatively involved chemical syntheses. Demand and production levels have risen steeply since then to their present estimated level of about 3300 t yr worldwide.

Physical and Chemical Properties

Salt Formation. As a weakly basic pyrimidine and a thiazolium cation, thiamine forms both mono- and dipositive salts, eg, the two commercial forms.

Thiamine chloride hydrochloride [67-03-8], (thiamine hydrochloride) $C_{12}H_{18}N_4OSCl_2$ (2), crystallizes as colorless monoclinic needles, mp 248–250°C (with decomposition), which in bulk appear white and have an approximate density of 0.4 kg L. Several polymorphic crystal forms have been reported. The salt has a characteristic thiazole meat-like odor and a slightly bitter taste. On exposure to air of average humidity, the hydrochloride (2) can adsorb up to one mole of water (more typically slightly less, to about 4% by weight), which may be removed by heating to 100°C or by vacuum drying. It is very soluble in water (over 1 kg L at 25°C), soluble in glycerol (0.056 kg L), propylene glycol, and methanol, sparingly soluble in 95% ethanol (0.01 kg L), and practically insoluble in less polar organic solvents. In 1–5% solutions in water it shows a pH of 3–3.5 (6,7).

Thiamine mononitrate [532-43-4] (3), $C_{12}H_{17}N_5O_4S$, is an apparently white, colorless crystalline solid with a typical odor, melting point of ca 196–200°C (dec.) and an approximate bulk density of 0.5 kg L. It is much less soluble in water than the hydrochloride (0.027 kg L at 25°C, 0.030 kg L at 100°C) and practically nonhygroscopic. Dilute solutions in water show a pH of 6.5–7 (6,8).

Numerous other salts have been reported in the literature, including some which are insoluble in water.

Physical Chemical Characterization. Thiamine, its derivatives, and its degradation products have been fully characterized by spectroscopic methods (9,10). The ultraviolet spectrum of thiamine shows pH-dependent maxima (11). 1H , ^{13}C , and ^{15}N nuclear magnetic resonance spectra show protonation occurs at the 1-nitrogen, and not the 4-amino position (12–14). The 1H spectrum in D_2O shows no resonance for the thiazole 2-hydrogen, as this is acidic and readily exchanged via formation of the thiazole ylid (13) an important intermediate in the biochemical functions of thiamine. Recent work has revised the pK_a values for the two ionization reactions to 4.8 and 18 respectively (9,10,15). The mass spectrum of thiamine hydrochloride shows no molecular ion under standard electron impact ionization conditions, but fast atom bombardment and chemical ionization allow observation of both an intense peak for the parent cation and its major fragmentation ion, the pyrimidinylmethyl cation (16).

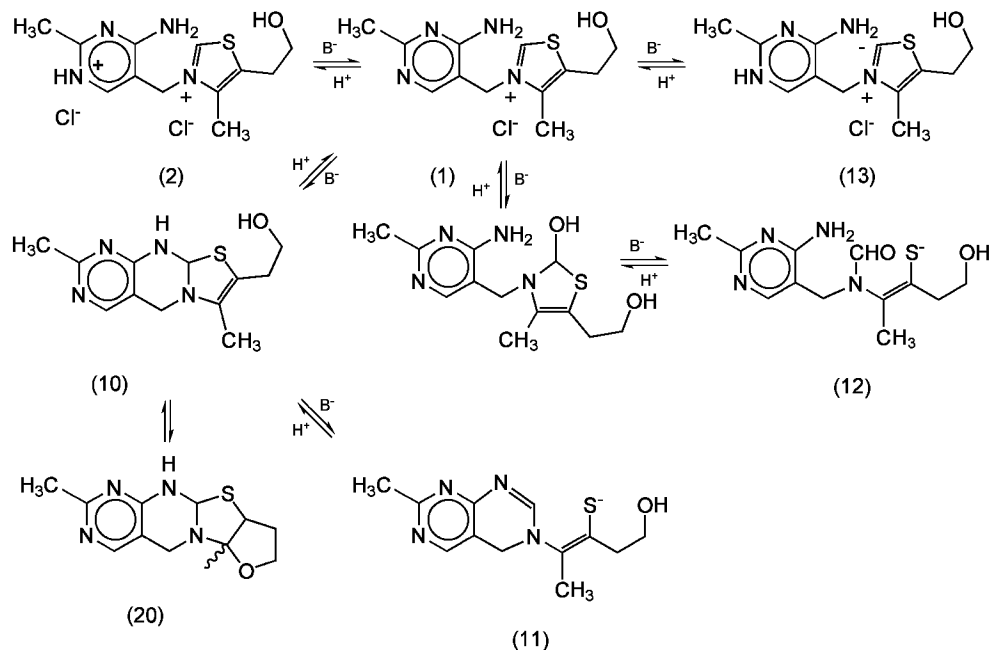
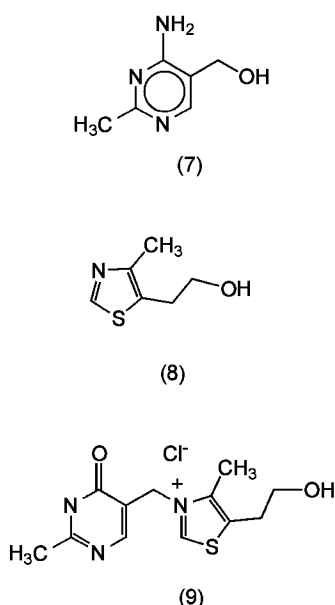


Fig. 1. Structural changes of thiamine with pH.

Reactions and Stability. Thiamine hydrochloride is stable as a solid if kept dry. Heating to 100°C for 24 h does not diminish its potency. In solution, its stability depends heavily on conditions of pH, temperature, and oxygen. Aqueous solutions below pH 5 are stable to oxygen, heating, and even autoclaving (11). Heating in water to 140°C under pressure gives decomposition to 4-amino-5-hydroxymethyl-2-methyl-pyrimidine [73-67-6] (7) and 4-methyl-5-(2-hydroxyethyl)thiazole [137-00-8] (8) (17). Heating with 20% hydrochloric acid hydrolyzes the 4-amino group to yield oxythiamine [582-36-5] (9), an antagonist of thiamine (18). Above pH 5, aqueous thiamine is destroyed relatively rapidly by boiling. At pH 7 destruction occurs even at room temperature. Concentrated solutions of thiamine in alcohol when neutralized degrade rapidly at room temperature to liberate thiazole (8) and form oligomers of the pyrimidine where the 5-methylene is linked to N-1 of another pyrimidine unit (10,14).



At pH 8 or above, even at room temperature, water solutions of thiamine turn yellow, then fade as a series of reactions ensues (9,10,19). Basification allows intramolecular attack of the 4-amino function on the thiazole ring to generate dihydrothiochrome [80483-97-2] (10), which eliminates thiolate to generate the salt of the yellow form, 5,6-dihydropyrimido(4,5*d*)pyrimidine [84825-03-6] (11), as a kinetic product (9,10,19) (Fig. 1). The sequence is reversible. More slowly, under the same conditions, a competing hydrolysis of the thiazolium ring via the pseudobase generates another product, known as thiamine thiol or the thiol form [554-45-0] (12) as the thermodynamic product. Also present in the system is a small amount of thiamine ylid [84812-92-0] (13), formed by deprotonation of C-2. The products dihydrothiochrome (10), yellow form (11), and ylid (13) undergo significant and useful chemistry.

The yellow form (11) on acidification is converted to the more stable thiol form (12). On oxidation, typically with alkaline ferricyanide, yellow form (11) is irreversibly converted to thiochrome [299-35-4] (14), a yellow crystalline compound found naturally in yeast but with no thiamine activity. In solution, thiochrome exhibits an intense blue fluorescence, a property used for the quantitative determination of thiamine.

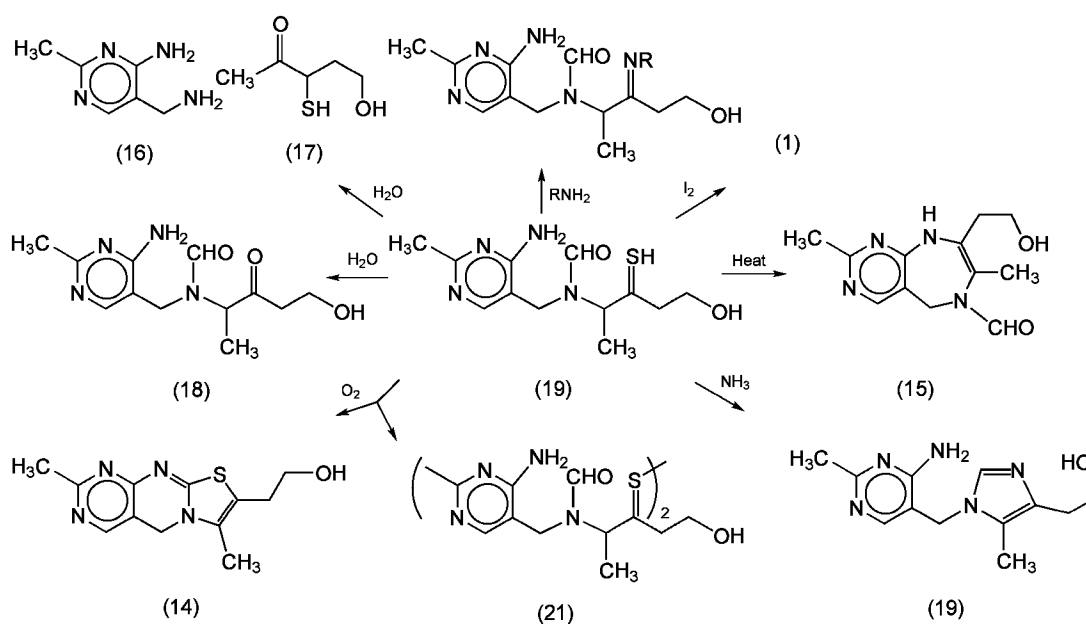


Fig. 2. Reactions of the thiol form (12).

The thiol form (12) undergoes reactions mainly via its *N*-formyl or ene-thiol groups. Heating an aqueous solution of the thiol form (Fig. 2) effects hydrolysis to the diamine 4-amino-5-aminomethyl-2-methyl-pyrimidine [95-02-3] (16), 5-hydroxy-3-mercaptopentan-2-one [15678-01-0] (17), and formic acid (20). Neutralization of a solution of the thiolate anion with carbon dioxide gives a fat-soluble basic material [21682-72-4, 35922-43-1] (20), presumably via dihydrothiochrome (10) (21). Acylation of the thiolate occurs on both sulfur and oxygen to give mono- or diacyl thiamines, some of which are interesting fat-soluble depot forms of thiamine.

The thiol form (12) is susceptible to oxidation (see Fig. 2). Iodine treatment regenerates thiamine in good yield. Heating an aqueous solution at pH 8 in air gives rise to thiamine disulfide [67-16-3] (21), thiochrome (14), and other products (22). The disulfide is readily reduced to thiamine *in vivo* and is as biologically active. Other mixed disulfides, of interest as fat-soluble forms, are formed from thiamine, possibly via oxidative coupling to the thiol form (12).

Whereas a claim of isolation of the thiamine ylid (13) has been the subject of controversy (15,23), kinetic and product studies and molecular orbital calculations support the formation and reactivity of a thiamine ylid as an unstable intermediate (15,24-26). Ylid (13) is accepted as an intermediate in explanations of the enzymatic and nonenzymatic reactions of thiamine. Among these reactions are the enzymatic oxidative decarboxylation of pyruvic and 2-ketoglutaric acids, the formation of acetoin, the reversible alpha-ketol transfer reactions catalyzed by transketolase, and the nonenzymatic acyloin condensation. According to the thiazolium ylid mechanism, ylid (13) reacts reversibly with carbonyl reagents, facilitating aldol-retroaldol and decarboxylation reactions (Fig. 3), via the acyl anion equivalent thiazolium ylids (23), the so-called active aldehydes, which are the key intermediates (27). In the specific case of oxidative decarboxylation of pyruvic acid, for example, ylid (23) transfers the acyl group to lipoic acid. Other reactions involving thiamine similarly involve nucleophilic additions of a thiazolium ylid to a suitable acceptor. Syntheses of a thiamine-pyruvate adduct (22) and a protonated form of its decarboxylation product (23) (28), as well as model studies with thiazolium ions (29), further support the thiazolium ylid mechanism. There is still current debate and investigation of the mechanistic details of the involvement of the thiamine ylid (13) in biological systems, including the role of the associated metal ions (15,30).

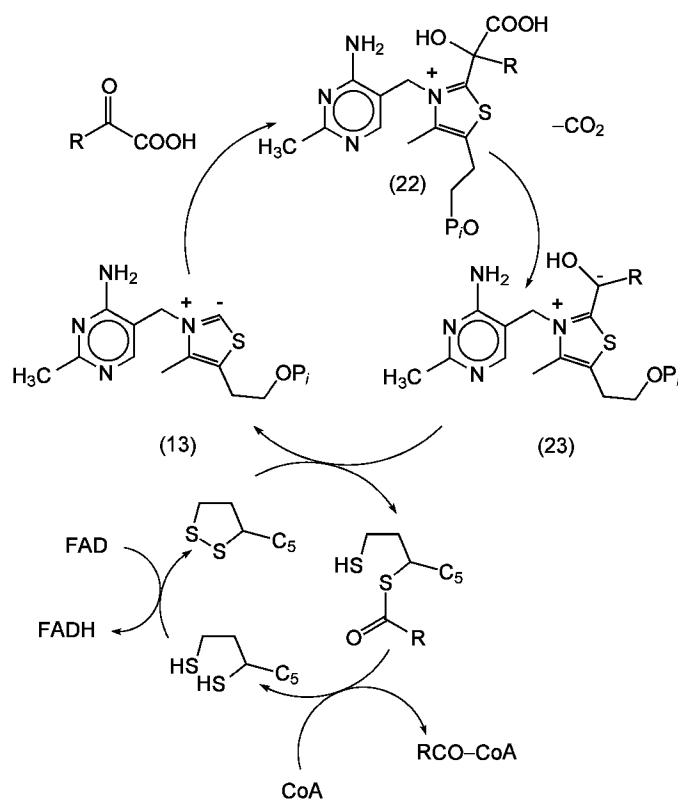


Fig. 3. Oxidative decarboxylation of alpha-ketoacids.

Thiamine is susceptible to reaction with various nucleophilic species. During the race for the isolation of thiamine, it was found by chance during attempted stabilization of extracts that thiamine is readily ruptured by sulfite treatment. This occurs slowly at pH 3 and rapidly at pH 5 and above, to give 4-amino-2-methyl-5-pyrimidine-methanesulfonic acid [108084-76-0] (24) (Fig. 4), with liberation of the thiazole moiety (8) (2,10). Other nucleophilic groups such as pyridines, phenolates, anilines, and azide react similarly in the required presence of sulfite. An addition-elimination mechanism involving sulfite attack on the pyrimidine ring has been elucidated (9,10). Similar reactions are observed in the degradation of thiamine by thiaminases, enzymes found in certain foods and bacteria, including some found in the human intestine (1,10,31). Thiaminase I (EC 2.5.1.2), found in shellfish, ferns, some vegetables and some bacteria, promotes the replacement of the thiazole by other organic bases, including purines. Thiaminase II (EC 3.5.99.2), found mostly in bacteria, catalyzes the cleavage of thiamine into (7), the biosynthetic precursor of thiamine and an antagonist of pyridoxine (vitamin B₆). A thiol group at the active site is strongly implicated (11). These enzymes can have strong effects on animals as they promote thiamine deficiency (32). Fortunately, they are thermolabile, and hence a problem mostly in uncooked foods or feeds.

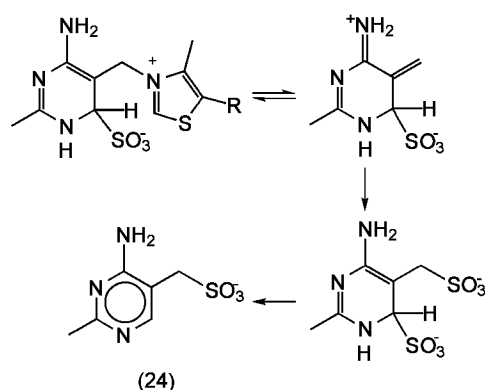


Fig. 4. Mechanism of nucleophilic degradation.

Thermostable substances which inactivate thiamine have been found in a large number of plants and in some animal tissues. Among these are polyphenols such as caffeic acid, tannic acid, hydroxylated derivatives of tyrosine, and some flavonoids (1). Thiamine is susceptible to destruction by x-rays, gamma rays, and ultraviolet light to generate diamine (16) by cleavage of the thiazole ring (1). Photochemical and thermal degradation of thiamine gives rise to numerous sulfur-containing heterocycles, some with meaty or bread aromas of interest to the flavor industry (33).

Operations in preparing or preserving food and feeds can sometimes lead to significant losses of thiamine value through leaching, chemical destruction (from high heat, high energy irradiation, or especially alkaline oxidizing conditions), and specific chemical interactions with thiamine-destroying substances. Such losses range in amounts from a few percent to as high as 85%. In many cases, the destruction of thiamine shows apparent first-order kinetics, and Arrhenius calculations can be used to estimate losses (1,34).

Thiamine forms the expected derivatives of the thiazole alcohol function, such as carboxylic and phosphate esters. Few reactions at the pyrimidine 4-amino function have been reported. Most of the usual conditions used for formation of amides, for example, lead to destruction of the thiazolium ring.

Natural Occurrence

Thiamine is widespread in nature, although generally in only relatively minute quantities (1,35). In microorganisms it is found mainly intracellularly, although minute amounts are lost to the natural environment upon cell lysis. In higher plants the most abundant form is free thiamine along with lesser amounts of the phosphate esters (4–6). Within the plant, thiamine is unevenly distributed, its amounts and location depending on the life stage. Seeds and uptake from the soil supply the plant until biosynthesis in the leaves begins. Thiamine is then transported from the leaves to the roots where it exhibits hormone-like effects on their growth. Later thiamine is again concentrated in the seeds, especially in the germ and in the pericarpal layers surrounding the starchy areas of the seed (36). As the germ and bran is often removed during processing for aesthetic reasons, highly refined rice or wheat, for example, can have

significantly lowered contents of thiamine, and can be a cause of thiamine deficiency when used as staple foods.

In the tissues of animals, most thiamine is found as its phosphorylated esters (4–6) and is predominantly bound to enzymes as the pyrophosphate (5), the active coenzyme form. As expected for a factor involved in carbohydrate metabolism, the highest concentrations are generally found in organs with high activity, such as the heart, kidney, liver, and brain. In humans this typically amounts to 1–8 µg g of wet tissue, with lesser amounts in the skeletal muscles (35). A typical healthy human body may contain about 30 mg of thiamine in all forms, about 40–50% of this being in the muscles owing to their bulk. Almost no excess is stored. Normal human blood contains about 90 ng mL, mostly in the red cells and leukocytes. A value below 40 ng mL is considered indicative of a possible deficiency. Amounts and proportions in the tissues of other animal species vary widely (31,35).

Good natural human dietary sources of thiamine are unrefined cereal grains, organ meats, pork, legumes, and nuts (Table 1). Oils, fats, and highly refined foods are essentially devoid of thiamine (3,36). Although thiamine is widespread in foods, some can be lost in food preparation and storage. As a result, dietary intake can vary significantly. In most developed countries, foods (typically white rice and white flour) and feeds are supplemented with thiamine and its use in vitamin tablets as dietary supplements is common. Enriched grains, cereals, and baked products contribute about 30–45% of the recommended daily allowance (RDA) for the adult diet in the United States.

Table 1. Average Thiamine Contents of Foods

Food	Thiamine, µg 100 g
wheat germ	2050
dried brewer's yeast	1820
soybeans	1300
pork	600–950
dried beans and peas	680
dried milk whey	500
nuts	300–560
brown rice	300
beef liver	300
potatoes	170
fish	50–90
eggs	70
vegetables (fruit, leaf, stem, root)	60–70
whole milk	30–70
white rice	50

Biochemical and Physiological Functions

Thiamine serves essential functions and its deficiency causes particularly deleterious effects on an organism's energy status and, in higher organisms, its nerve functions. In living systems, the only established biochemically active form is the pyrophosphate (cocarboxylase) (5), which plays a vital role in intermediate metabolism as a cofactor for some important enzymatic reactions. Dehydrogenase enzymes require cocarboxylase (5) for oxidative decarboxylation of 2-ketoacids, notably pyruvate in glycolysis, 2-oxoglutarate in the citric acid cycle, and other ketoacids from amino acid decarboxylation (1). Transketolase enzymes require cocarboxylase (5) for the reversible transfer of alpha-ketols in ketose–aldose transformations important in the production of pentoses for RNA and DNA synthesis (1). Thiamine triphosphate (6) occurs in unusually higher concentrations in nerve tissues and the brain and may play an essential role in the stimulation of peripheral nerves (35).

In humans, thiamine is both actively and passively absorbed to a limited level in the intestines, is transported as the free vitamin, is then taken up in actively metabolizing tissues, and is converted to the phosphate esters via ubiquitous thiamine kinases. During thiamine deficiency all tissues stores are readily mobilized. Because depletion of thiamine levels in erythrocytes parallels that of other tissues, erythrocyte thiamine levels are used to quantitate severity of the deficiency. As deficiency progresses, thiamine becomes undetectable in the urine, the primary excretory route for this vitamin and its metabolites. Six major metabolites, of more than 20 total, have been characterized from human urine, including thiamine fragments (7,8), and the corresponding carboxylic acids (1,37,38).

The classic pathology resulting from severe thiamine deficiency in humans is called beriberi. Similar conditions have been described for many other animals. Beriberi develops primarily from inadequate intake of thiamine or from ingestion of food containing antithiamine factors and is somewhat rare in developed areas of the world. Less severe thiamine deficiency is more common and is characterized by anorexia and mental disturbances, such as irritability, inattention, memory defects, depression, and insomnia. If a lesser deficiency is left untreated, one of several clinical forms of beriberi develops, the symptoms of which include mental changes, peripheral neuritis, paresthesias, muscle cramps, edema, muscular atrophy, and cardiac failure. The most commonly encountered type of thiamine deficiency in Western countries is associated with alcohol abuse. This is generally thought to result from high intake of empty calories and low intake of nutritionally adequate foods. Other factors that influence thiamine status include general level of muscular activity; dietary practices such as high intake of refined carbohydrates, tea or coffee, or raw seafood; reduced thiamine intake as a result of disease, or parasites or drugs lowering food intake and utilization or thiamine absorption; pregnancy and lactation; heavy smoking; advanced age; genetic background; and stress. In other animals, climate and intestinal microflora also become important, and toxic effects can occur from changes in the microflora (39). Thiamine status has been monitored by blood thiamine levels, thiamine and metabolite levels in the urine, blood pyruvate and lactate levels, and blood transketolase activity (1,37,38).

Thiamine requirements vary and, with a lack of significant storage capability, a constant intake is needed or deficiency can occur relatively quickly. Human recommended daily allowances (RDAs) in the United States are based on calorie intake at the level of 0.50 mg 4184 kJ (1000 kcal) for healthy individuals (Table 2). As little as 0.15–0.20 mg 4184 kJ will prevent deficiency signs but 0.35–0.40 mg 4184 kJ are required to maintain near normal urinary excretion levels and associated enzyme activities. Pregnant and lactating women require higher levels of supplementation. Other countries have set different recommended levels (1,37,38).

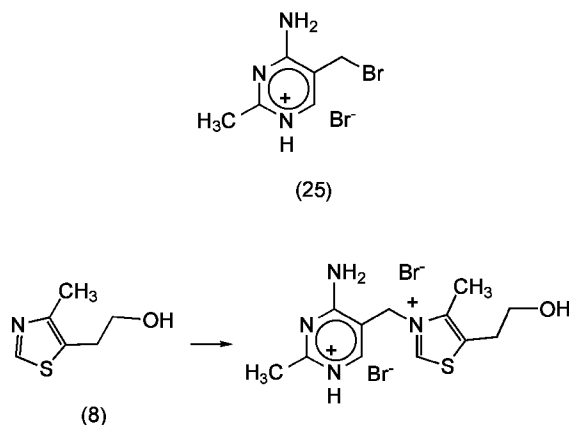
Table 2. U.S. RDAs for Thiamine

Population segment	RDA, mg
infants and children <4 yr	0.7
adults and children >4 yr	1.5
pregnant or lactating women	1.7

Manufacture

Chemistry. Isolation of thiamine from natural sources (rice bran, yeast extracts, or wheat germ) is only of historical interest. Production by bioprocesses is not cost-effective at present. All of the thiamine produced worldwide is manufactured by chemical processes operated at moderately large scale. Two major synthetic routes have been used: alkylation of a preformed thiazole, or construction of the thiazolium salt from a pyrimidine carrying the ultimate thiazole nitrogen (9,40). The latter approach is now generally preferred for manufacturing.

The first approach parallels the known biosynthetic pathway where the alkylating agent is the pyrophosphate ester of alcohol (7). Typical of this approach is Williams' first convergent synthesis, which is the basis for the first industrial method developed by Merck & Co. (2,5,41). Synthetic 4-amino-5bromomethyl-2-methylpyrimidine [2908-71-6] (25) and thiazole (8), or its *O*-acetate, were condensed and then the bromide was replaced by use of silver salts. Numerous variants were published or patented in the 1930s and 1940s. Such methods generally gave only moderate yields, used higher temperatures in polar solvents, required tedious purification of the colored products and are no longer used.



In the second general method, the amine (16), known as Grewe diamine, is the paradigm intermediate onto which the thiazolium ring is constructed. Differences occur only in the raw materials and methods used for the manufacture of the diamine. The same end of the linear sequence is universally practiced by all large manufacturers.

The synthesis which forms the basis of production at Hoffmann-La Roche (Fig. 5) proceeds via the pyrimidinenitrile [698-29-3] (26) made from malononitrile, trimethylorthoformate, ammonia, and acetonitrile (42,43). High pressure catalytic reduction of the nitrile furnishes diamine (16). The overall sequence is short, highly efficient, and generates almost no waste. However, malononitrile is a relatively expensive and hazardous three-carbon source.

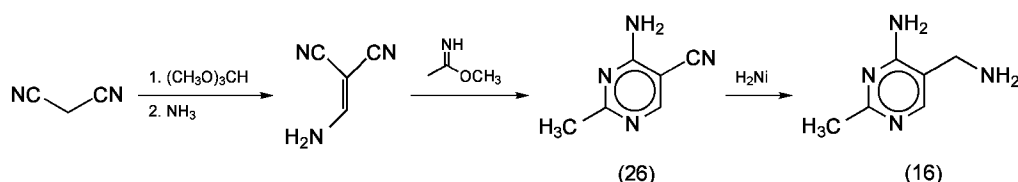


Fig. 5. Roche process.

Other syntheses produce Grewe diamine using inexpensive acrylonitrile and alkyl formates as raw materials. Such routes deliver the pyrimidine bridge carbon at the correct aminomethyl oxidation level without a need for reduction. Thus, in the older method of Shionogi, base-catalyzed acylation of 3-methoxypropanenitrile followed by enolate alkylation and ring closure with two equivalents of acetamidine gives the acetyl derivative [23676-63-3] (27) of Grewe diamine (Fig. 6) (44–46). Alternatively, in a newer method by BASF, acylation of 3-formamidopropanenitrile, followed by ring closure with acetamidine, gives diamine as its formamide [1886-34-6] (28) (Fig. 7) (47–50). Both the Shionogi and BASF methods are simple technologies based on inexpensive raw materials, but both generate significant levels of salts and organic carbon as wastes.

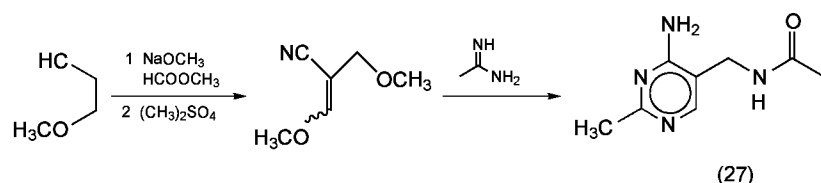


Fig. 6. Shionogi process.

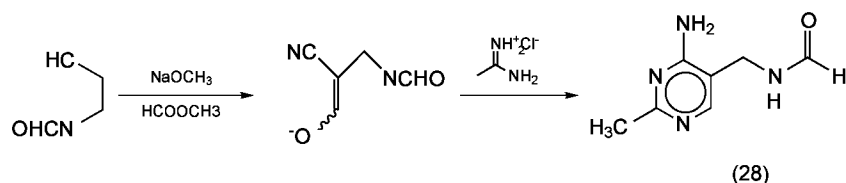


Fig. 7. BASF process.

In the past decade Takeda and Ube, in a joint venture have developed a new process for diamine, also based on acrylonitrile and alkyl formate, but carrying the pyrimidine bridge carbon at the carbonyl level (Fig. 8) (50–55). Metal-catalyzed oxidation of acrylonitrile in methanol generates 3,3-dimethoxypropanenitrile [57597-62-3] (29) which is acylated with methoxide carbon monoxide under pressure. Unlike other methods in which the intermediate enolate is alkylated, in this process the enolate is acetalized and the acetal thermally converted to the acetal enol ether [87466-78-2] (30). Condensation with acetamidine provides the remaining carbons. After hydrolysis of the acetal [16057-06-0] (31), reductive amination at high pressure with a specialized catalyst provides diamine (16) in very high overall yield. The process is operated in a new, large-scale, automated, continuous, technically complex plant in Japan. Advantages include use of inexpensive acrylonitrile and carbon monoxide, avoidance of the costs of an alkylating agent, and consumption of only one equivalent of acetamidine.

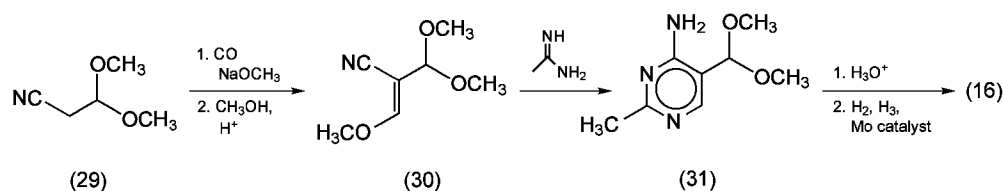
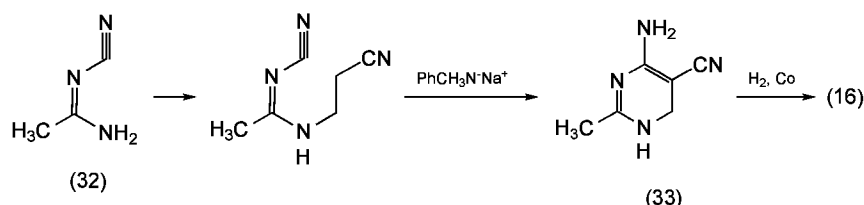


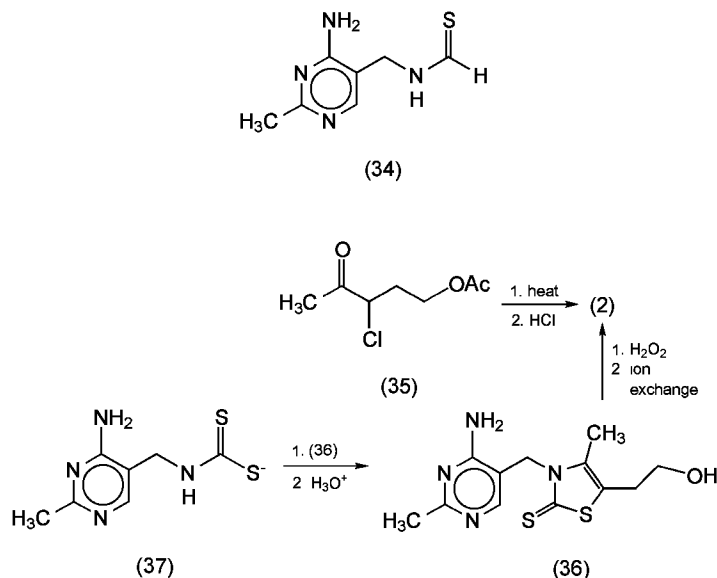
Fig. 8. Takeda-Ube process.

In a much different approach based on cyanamide, acrylonitrile, and acetonitrile, cyanoacetamidine [56563-07-6] (32) is cyanoethylated and the condensation product [56563-10-1] (33) is dehydrogenated and hydrogenated directly to Grewe diamine in the presence of Raney cobalt and ammonia (56).

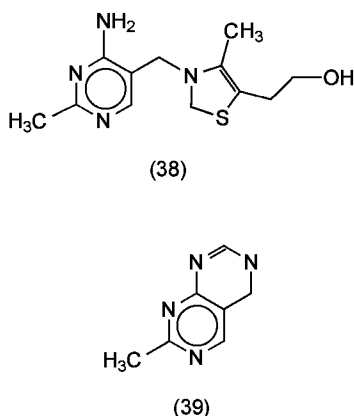


Methods for synthesis of the thiazolium ring also have matured technically based on the cost, throughput, and waste disposal issues of production. In earlier syntheses, the 2-carbon and the sulfur atom were supplied as potassium dithioformate, made from chloroform and potassium sulfide.

N-(4-Amino-2-methyl-5-pyrimidinyl)-methylthioformamide [31375-20-9] (34) served as a partner to 5-acetoxy-3-chloropentan-2-one [119867-66-2] (35) in a classical Hantzsch thiazole synthesis (3,57). Such routes suffered from similar limitations to the quaternization routes mentioned before, but were used in commercial production until the discovery of the paradigm intermediate known as thiothiamine [299-35-4] (36). In the 1940s and 1950s, chemists at Tanabe and Takeda showed that the 2-carbon and the sulfur atom can be supplied very efficiently via carbon disulfide, the extra sulfur atom being readily removed by oxidation in nearly quantitative yield (58,59). Typically, an aqueous solution of diamine and alkali is treated in succession with carbon disulfide to form the dithiocarbamate [2882-49-7] (37), then with chloroketone (35), then acid to form the relatively insoluble, thiothiamine (36). Oxidation with hydrogen peroxide forms thiamine sulfate, which is converted by ion exchange (qv) to a solution of the hydrochloride (2) which is concentrated, crystallized, and dried. The much less soluble nitrate (3) is precipitated from aqueous solution with alkali metal nitrate. The advantages of this approach, ie, a cheap, easily handled sulfur source, very high overall yield and throughput, excellent product color and purity, and the use of water as solvent, have made it the preferred method of manufacture worldwide. Disadvantages include use of chlorinated intermediates, salt load from the sulfate waste stream, and malodorous aqueous waste which must be well contained and treated. This chemistry is practiced by Takeda and Hoffmann-La Roche in automated factories using standard glass and stainless steel fine chemical processing equipment operated continuously at many hundreds of metric tons annually. Production in the several smaller Chinese factories is probably basically similar.

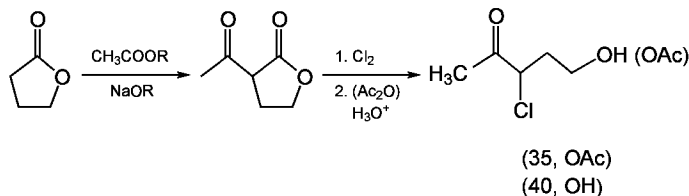


In two unique, convergent approaches to the thiazole ring, other formate synthons are used for the 2-carbon. In one, reaction of formaldehyde with diamine (16) and thiol (17) gives dihydrothiamine [959-18-2] (38), which is oxidized to thiamine (60). In another, reaction of diamine (16) with orthoformate gives intermediate dihydropyrimidine [31375-19-6] (39), to which thiol (17) is added. Acidic rearrangement gives thiamine in high yield (60).



Although numerous other materials have been proposed as carbon sources for the thiazole ring, all manufacturing of thiamine is believed to use

3-chloro-5-acetoxypentan-2-one (35) or the corresponding alcohol [13045-13-1] (40) as intermediates. These are made by chlorination of acetylbutyrolactone, the latter from inexpensive butyrolactone and methyl acetate, generating chlorinated wastes.



Worldwide production of thiamine was estimated at 3300 t in 1995. The principal suppliers were Hoffmann-La Roche, Takeda, and several Chinese factories. Prices in the United States were in the range of \$20–\$28 kg in 1995.

Product Specifications and Testing. Nutrients and diet supplements without claims are considered foods, and thus are regulated by the U.S. Food and Drug Administration and are further subject to specific food regulations. Specifications for the hydrochloride (2) and the mononitrate (3) for foods are given in the *Food Chemicals Codex* (62) and for pharmaceuticals in the *U.S. Pharmacopeia* (63). General test methods have been summarized (64).

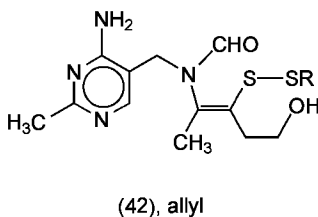
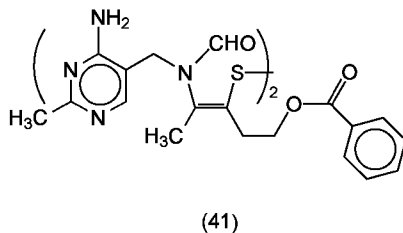
Safety and Handling. The hydrochloride (2) and the nitrate (3) are typically packaged in standard 20–25-kg foil or poly-lined cardboard boxes. Smaller packaging sizes are also available. Both hydrochloride (2) and nitrate (3) are required to be kept under normal cool, dry storage conditions, not to exceed 70°C and protected from light. Both are listed in the TSCA Inventory and material safety data sheets (MSDS) are available from suppliers. Precautions against dust exposure to the eyes, skin, or lungs (IOEL of 3.0 mg m³ time weighted average) and against fire and dust explosion are indicated. The NPFA ratings for health, fire, and reactivity for the hydrochloride (2) are 1, 2, and 1, respectively, and for the nitrate (3) they are 1, 3, and 1. Neither national nor international transportation is regulated. The U.S. EPA has not established reportable quantities for environmental releases (7,8). The acute LD₅₀ (rat, oral) of 12,300 mg kg for the hydrochloride (2) and 15,900 for the nitrate (3) qualify these materials as relatively harmless orally. There is no evidence of toxicity from thiamine taken orally by humans, even when doses of 500 mg, over 300 times the RDA, are taken daily for a month. Excess thiamine is easily cleared by the kidney (7,8,65). Parenteral doses of the hydrochloride (subcutaneous, intramuscular, or intravenous) have generally been well tolerated up to 100–500 mg with very few toxic effects. Some cases of sensitization or anaphylactic shock on repeated injections have been reported (1).

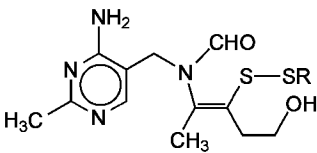
Analytical Methods

Fluorometric, chromatographic, microbiological, and animal assays have been used for thiamine and its derivatives (66). The most widely used and officially sanctioned method has been the fluorometric assay (67), although high performance liquid chromatography has been increasingly employed (68). In natural materials, thiamine is often present as its phosphate esters and is protein-bound, therefore procedures to free it are necessary steps in most assays. Determination of thiamine by the fluorometric method involves acid extraction, phosphatase hydrolysis, ion exchange to remove interferences, and oxidation with ferricyanide or other reagents to thiochrome, whose fluorescence (excitation 365 nm emission 435 nm) is measured and compared with a standard (67). Various hplc methods have been developed for quantitation of thiamine and its phosphates in biological matrices to picomole to femtomole levels. Separations are made with reversed-phase, ion-exchange, and specialized straight-phase packings and detection with uv, post-column derivatization—fluorometry, or electrochemical techniques. Alternatively, precolumn derivatization and measurement of thiochrome has been used (68). High performance capillary electrophoresis has been applied (69). Gas chromatography has been used to determine thiamine to picomole level as the thiazole portion following cleavage of thiamine with sulfite (71). Microbiological assays are simple, inexpensive, and sensitive (5–50 ng thiamine). The results agree well with fluorometric methods but can take longer to achieve results. Microorganisms that have been most widely used include *Lactobacillus fermenti* (ATCC 9338) and *Lactobacillus viridescens* (ATCC 12706), partly because they respond only to intact thiamine and not to its precursors (31). Bioassays in rats based on growth or cure of polyneuritis symptoms are time consuming and expensive but have been historically valuable in determining thiamine availability in foods and thiamine activity of related compounds (31).

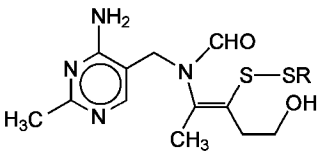
Forms, Derivatives, and Analogues

The hydrochloride and mononitrate are the only commercial forms approved in the United States. The mono-, di-, and triphosphate esters (4–6) are colorless, water-soluble, organic-insoluble, high melting solids found naturally. Although they have been synthesized (9), they have not been produced commercially on large scale. In Japan and Europe, other forms have been approved for human use. Thiamine disulfide (21) is somewhat more soluble than thiamine in organic media. This material and its more fat-soluble *O,O*-dibutyrate and *O,O*-dibenzoate [2667-89-2] (41) have been used therapeutically for treatment of thiamine deficiency (70). Treatment of thiamine with extracts of garlic or other *Allium* species converts it to a lipid-soluble disulfide derivative which is a very physiologically active depot form of thiamine. Thiamine allyl disulfide [554-44-9] (allithiamine) (42), was the first of several thiamine alkyl disulfides to be used therapeutically for thiamine deficiency, including prosultiamine [59-58-5] (43) and fursultiamine [804-30-8] (44). Because of their greater lipid solubility, they are absorbed and retained more strongly in the body. *In vivo* they are converted back to thiamine. They are of interest almost exclusively in Japan and are not yet approved in the United States (70,71). Sugar derivatives (qv) are claimed to have better taste and no odor (73). In applications where water solubility is detrimental, such as fish feeds with high thiaminase activity, formulations of the hydrochloride or nitrate with a waxy coating are effective.



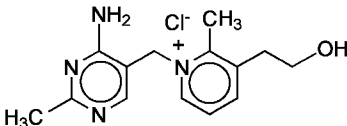


(43), propyl

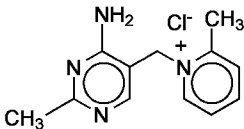


(44), tetrahydrofurfuryl

Numerous analogues of thiamine have been synthesized by structural modifications of the pyrimidine or thiazole rings or the bridging atoms (1,9,10). None has been found to exceed the biological activity of thiamine but some act as antagonists. Substitution at the 2-methyl, 4-amino, or 6-position of carbon for ring nitrogens, or of the methylene bridge, results in loss of activity. In the thiazolium ring, the sulfur atom, hydrogen at the 2-position, and the 2-hydroxyethyl group are essential for activity. The pyridine analogue, pyriothiamine [534-64-5] (45) is the most potent thiamine antagonist known. Some thiamine antagonists were found at low levels to selectively inhibit thiamine uptake by *Coccidia*. A modified thiamine-like molecule, called Amprolium [121-25-5] (46) was once claimed to be useful as a coccidiostat for poultry (73).



(45)



(46)

Biosynthesis

Higher plants, most bacteria, and some fungi biosynthesize thiamine. Humans or most other animals cannot, although some of their gut microflora can. Many microbial species are self-sufficient, others can synthesize thiamine if one or both of immediate precursors are available, and still others require the complete substance. Few produce much of an excess over their own requirements. Biosynthesis of thiamine in microorganisms has been reviewed (9,74–76). Media modifications and mutagenesis have achieved small increases of thiamine levels with microorganisms, the highest levels of a few mg L being found with yeasts (qv) (77). A broader search among many yeast species has reported a *Saccharomyces cerevisiae* strain which accumulates up to 200 mg L in the culture broth (78).

The pathways for thiamine biosynthesis have been elucidated only partly. Thiamine pyrophosphate is made universally from the precursors 4-amino-5-hydroxymethyl-2-methylpyrimidine pyrophosphate [841-01-0] (47) and 4-methyl-5-(2-hydroxyethyl)thiazole phosphate [3269-79-2] (48), but there appear to be different pathways in the earlier steps. In bacteria, the early steps of the pyrimidine biosynthesis are same as those of purine nucleotide biosynthesis, 5-Aminoimidazole ribotide [41535-66-4] (AIR) (49) appears to be the sole and last common intermediate; ultimately the elements are supplied by glycine, formate, and ribose. AIR is rearranged in a complex manner to the pyrimidine by an as-yet undetermined mechanism. In yeasts, the pathway to the pyrimidine is less well understood and may be different (74–83) (Fig. 9).

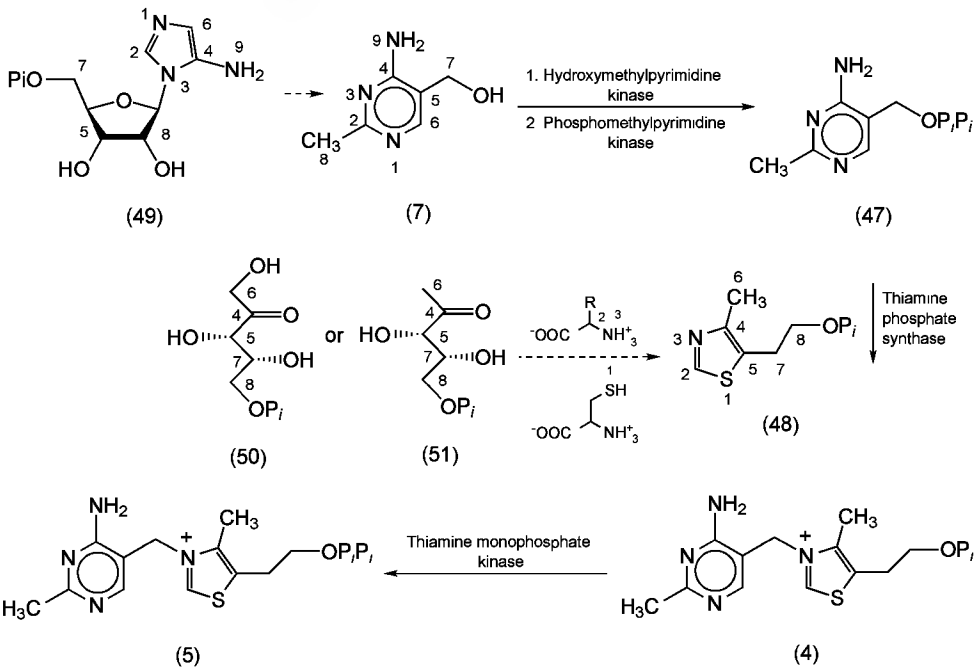


Fig. 9. Biosynthesis of thiamine.

In the biosynthesis of the thiazole, cysteine is the common sulfur donor. In yeasts, the C-2 and N may be supplied by glycine, and the remaining carbons by D-ribulose-5-phosphate [108321-99-9] (50). In anaerobic bacteria, the C-2 and N may be recruited from tyrosine and the carbons from D-1-deoxyxylulose [16709-34-5] (51), whereas in aerobic bacteria the C-2 and N may be derived from glycine, as in yeasts 7 (74–76,83–86) (see Fig. 9). Biosynthesis of pyrophosphate (5) from pyrimidine phosphate (47) and thiazole phosphate (48) depends on the activity of five enzymes, four of them kinases (87). In yeasts and many other organisms, including humans, pyrophosphate (5) can be obtained from exogenous thiamine in a single step catalyzed by thiamine pyrophosphokinase (88). A number of the genes involved in the biosynthesis of thiamine in *E. coli* (89–92), *Rhizobium meliloti* (93), *B. subtilis* (94), and *Schizosaccharomyces pombe* (95,96) have been mapped, cloned, sequenced, and associated with biosynthetic functions. Thiamine biosynthesis is tightly controlled by feedback and repression mechanisms limiting overproduction (97,98). A cost-effective bioprocess for production of thiamine will require significant additional progress.

Uses

Most of the thiamine sold worldwide is used for dietary supplements. Primary market areas include the following applications: addition to feed formulations, eg, poultry, pigs, cattle, and fish (see FEEDS AND FEED ADDITIVES); fortification of refined foods, eg, flours, rice, and cereal products; and incorporation into multivitamins. Small amounts are used in medicine to treat deficiency diseases and other conditions, in agriculture as an additive to fertilizers (qv), and in foods as flavorings. Generally for dry formulations, the less soluble, nonhygroscopic nitrate is preferred. Only the hydrochloride can be used for intravenous purposes. Coated thiamine is used where flavor is a factor.

BIBLIOGRAPHY

"Thiamine" in *ECT* 1st ed., Vol. 14, pp. 38–48, by E. Pierson, Merck & Co., Inc; "Thiamine" under "Vitamins" in *ECT* 2nd ed., Vol. 20, pp. 173–184, by P. I. Pollak, Merck & Co., Inc.; in *ECT* 3rd ed., Vol. 24, pp. 124–139, by Y. Oka, Takeda Chemical Industries, Ltd.

1. C. J. Gubler, in L. J. Machlin, ed., *Handbook of Vitamins*, 2nd ed., Marcel Dekker, New York, 1991, pp. 233–280.
2. R. R. Williams and J. K. Cline, *J. Am. Chem. Soc.* **58**, 1504 (1936); *ibid.*, **59**, 1052 (1937).
3. A. R. Todd and F. Bergel, *J. Chem. Soc.*, 364 (1936).
4. A. R. Todd and F. Bergel, *J. Chem. Soc.*, 26 (1938).
5. H. Andersag and K. Westphal, *Ber. Dtsch. Chem. Ges.* **70**, 2035 (1937).
6. *The Merck Index*, 11th ed., Merck & Co., Rahway, N.J., 1989, p. 1464.
7. *Thiamine Hydrochloride*, MSDS 3834, Hoffmann-La Roche Inc., Nutley, N.J.
8. *Thiamine Mononitrate*, MSDS 3835, Hoffmann-La Roche Inc., Nutley, N.J.
9. O. Isler, G. Brubacher, S. Ghisla, and B. Kraeutler, *Vitamine II*, Georg Thieme Verlag, New York, 1988, pp. 17–49.
10. J. A. Zoltewicz and G. Uray, *Bioorganic Chem.* **22**, 1–28 (1994) and references therein.
11. J. M. E. H. Chahine and J. E. Dubois, *J. Am. Chem. Soc.* **105**, 2335 (1983).
12. J. D. Roberts and co-workers, *J. Am. Chem. Soc.* **99**, 6423 (1977).
13. A. Lycka and J. Cizmarik, *Pharmazie* **45**, 371 (1990).
14. J. M. E. H. Chahine, *J. Chem. Soc., Perkin Trans. II*, 505 (1990).
15. R. Kluger, *Chem. Rev.* **87**, 863 (1987).
16. R. D. Sedgwick and co-workers, *J. Chem. Soc., Chem. Commun.*, 325 (1987).
17. J. J. Windheuser and T. Higuchi, *J. Pharm. Sci.* **51**, 3545 (1962).
18. T. Matsukawa and S. Yurugi, *Yakugaku Zasshi* **71**, 827 (1951).
19. J. Herrmann and co-workers, *J. Chem. Soc., Perkin Trans. II*, 463 (1995).
20. A. Watanabe and Y. Asahi, *Yakugaku Zasshi* **77**, 153 (1957).
21. A. Takamizawa and co-workers, *Chem. Pharm. Bull.* **19**, 759 (1971).
22. T. Matsukawa and S. Yurugi, *Yakugaku Zasshi*, **72**, 1599 (1952).
23. H. Sugimoto, K. Hirai, and co-workers, *J. Org. Chem.* **55**, 46 (1990).
24. J. Fournier, *Bull. Soc. Chim. Fr.*, 854 (1988).
25. J. M. E. H. Chahine and J. E. Dubois, *J. Chem. Soc., Perkin Trans. II*, 25 (1989).
26. Y.-T. Chen and F. Jordan, *J. Org. Chem.* **56**, 5029 (1988).
27. R. Breslow, *J. Am. Chem. Soc.* **80**, 3719 (1958).
28. J. T. Stiners and M. W. Washabaugh, *Bioorg. Chem.* **18**, 425 (1990).
29. R. Breslow, *Pure Appl. Chem.* **62**, 1859–1866 (1990).
30. M. Louloudi and N. Hadjiliadis, *Coord. Chem. Rev.* **135/136**, 429–468 (1994).
31. W. H. Sebrell and R. S. Harris, eds., *The Vitamins*, 2nd ed., Vol. V, Academic Press, Inc., New York, 1972, p. 98.
32. W. C. Evans, in P. L. Munson, J. Glover, E. Diezfalusy, and R. E. Olson, eds., *Vitamins and Hormones*, Vol. 33, Academic Press, Inc., New York, 1975, pp. 467–504.
33. P. Werkhoff and co-workers, *ACS Symp. Ser.* **564**, 199–223 (1994).
34. L. Mauri and co-workers, *Int. J. Food Sci. Technol.* **24**, 1–9 (1989).
35. M. P. Lamden, in W. H. Sebrell, Jr. and R. S. Harris, eds., *The Vitamins*, 2nd ed., Vol. V, Academic Press, Inc., New York, 1972, pp. 114–120.
36. A. Mozafar, *Plant Soil* **167**, 305–311 (1994).
37. R. E. Davis and G. C. Icke, *Adv. Clin. Chem.* **23**, 93 (1983).
38. P. M. Finglas, *Int. J. Vit. Nutr. Res.* **63**, 270 (1993).
39. J. Harmeyer and U. Kollenkirchen, *Nutr. Res. Rev.* **2**, 201–225 (1989).
40. D. Brown and co-workers, *The Pyrimidines*, 2nd ed., John Wiley & Sons, Inc., New York, 1994.
41. U.S. Pat. **2,235,862** (Mar. 25, 1940), O. Zima (to Merck & Co.).
42. R. Grewe, *Z. Physiol. Chem.* **242**, 89 (1936).
43. Fr. Pat. 831,110 (Aug. 23, 1938), (to Hoffmann-La Roche).
44. A. Takimizawa and K. Hirai, *Chem. Pharm. Bull.* **12**, 393 (1964).
45. H. Miromoto and co-workers, *Chem. Ber.* **106**, 893 (1973).
46. K. Tokuyama and co-workers, *Bull. Chem. Soc. Jap.* **46**, 253 (1973).
47. U.S. Pat. **4,226,799** (Oct. 16, 1978), W. Bewert and W. Littmann (to BASF).
48. Ger. Pat. **3,511,273** (Mar. 28, 1985), W. Ernst and J. Paust (to BASF).
49. Ger. Pat. **3,520,982** (Dec. 18, 1986), H. Kiefer and W. Bewert (to BASF).
50. K. Nishihara and co-workers, *Kagaku Kagaku* **55**, 433 (1991).
51. Eur. Pat. **55,108** (June 30, 1982), K. Matsui and co-workers (to Ube Industries).

52. **Eur. Pat. 279,556** (Aug. 24, 1988), K. Nishihara and co-workers (to Ube Industries).

53. **U.S. Pat. 4,536,577** (Aug. 20, 1985), H. Yoshida and S. Niida (to Ube Industries).

54. **Ger. Pat. 3,303,789** (Aug. 11, 1983), H. Yoshida and co-workers (to Ube Industries).

55. **Rus. Pat. 6,307,869** (July 10, 1988), Y. Miyashiro and co-workers (to Takeda Chemical Industries).

56. A. Edenhoffer and co-workers, *Helv. Chim. Acta* **58**, 1230 (1975).

57. T. Matsukawa and co-workers, *J. Pharm. Soc. Japan* **63**, 216 (1943).

58. K. Washimi, *J. Pharm. Soc. Japan* **66**, 62 (1946).

59. T. Matsukawa and S. Hojiro, *J. Pharm. Soc. Japan* **69**, 550 (1949); *ibid.*, **70**, 28 (1950); **71**, 667, 720, 1215 (1951).

60. H. Hirano and Y. Yonemitsu, *J. Pharm. Soc. Japan* **76**, 1332 (1956).

61. G. Moine and co-workers, *Helv. Chim. Acta* **73**, 1300 (1990).

62. Food and Nutrition Board, National Research Council, *Food Chemicals Codex*, 3rd ed., National Academy Press, Washington, D.C., 1981, pp. 324–325.

63. *The United States Pharmacopeia XXII* (USP XXII-NF XVII) United States Pharmacopeial Convention, Inc., Rockville, Md., 1990, pp. 1356, 1357–1358.

64. F. J. Al-Shammary and co-workers, *Anal. Profiles Drug Subst.* **18**, 414 (1989).

65. H. E. Samerlich and L. J. Machlin, *Ann. NY Acad. Sci.* **699**, 1 (1992).

66. W. C. Ellefson, in J. Augustine, B. P. Klein, D. A. Becker, and P. B. Venugopal, eds., *Methods of Vitamin Assay*, 4th ed., John Wiley & Sons, Inc., New York, 1985, pp. 351–352.

67. K. Helrich, ed., *Official Methods of Analysis of the Association of Official Analytical Chemists*, 15th ed., Association of Official Analytical Chemists, Inc., Arlington, Va., 1990, pp. 942.23, 953.17, 957.17, and 986.27.

68. T. Kawasaki, in A. P. Leenheer, W. E. Lambert, and M. G. M. DeRuyter, eds., *Modern Chromatographic Analysis of the Vitamins*, Marcel Dekker, New York, 1992, pp. 319–354.

69. U. Jegle, *J. Chrom.* **652**, 495–501 (1993).

70. C. Kawasaki, in R. Hamb, I. Wool, and J. Lomaine, eds., *Vitamins and Hormones*, Vol. 21, Academic Press, Orlando, Fla., 1963, pp. 69–111.

71. R. E. Echols and co-workers, *J. Chrom.* **347**, 89–97 (1985).

72. C. E. Mueller and co-workers, *Int. J. Pharm.* **57**, 41–47 (1989).

73. Y. Susuki and K. Uchida, *Biosci. Biotech. Biochem.* **58**, 1273 (1994).

74. W. H. Ott and co-workers, *Poult. Sci.* **44**, 920 (1965).

75. A. Iwashima in E. Vandamme, ed., *Biotechnology of Vitamins, Pigments and Growth Factors*, Elsevier, Amsterdam, the Netherlands, 1988.

76. D. W. Young, *Nat. Prod. Rep.* **3**, 395 (1986).

77. G. M. Brown and J. M. Williamson, in F. C. Neidhardt, ed., *Escherichia coli and salmonella typhimurium*, American Society for Microbiology, Washington, D.C., 1987, p. 521.

78. A. Silhamkeva, *J. Inst. Brew.* **191**, 78 (1985).

79. G. G. Stewart and co-workers, *Can. J. Microbiol.* **38**, 1156 (1992).

80. T. Kozlok and I. D. Spenser, *J. Am. Chem. Soc.* **109**, 4968 (1987).

81. D. M. Downs and L. Peterson, *J. Bacteriol.* **176**, 4858 (1994).

82. N. J. Leonard and co-workers, *Proc. Natl. Acad. Sci.* **85**, 7174 (1988).

83. B. Estramareix and co-workers, *Biochem. Biophys. Acta* **1035**, 154 (1990).

84. K. Tazuya and co-workers, *Biochem. Mol. Biol. Int.* **30**, 893 (1993).

85. K. Tazuya and co-workers, *Biochem. Int.* **10**, 689 (1985).

86. R. H. White and F. B. Rudolph, *Biochem. Biophys. Acta* **542**, 340 (1978).

87. K. Tazuya and co-workers, *Biochem. Int.* **14**, 153 (1987).

88. Y. Kawasaki, *J. Bacteriol.* **175**, 5153 (1993).

89. A. Iwashima and co-workers, *J. Biol. Chem.* **268**, 17440 (1993).

90. T. Nohno, Y. Kasai, and T. Saito, *J. Bacteriol.* **170**, 4097 (1988).

91. N. Imamura and H. Nakayama, *J. Bacteriol.* **151**, 708 (1982).

92. J. Ryals and co-workers, *J. Bacteriol.* **151**, 899 (1982).

93. T. P. Begley and co-workers, *J. Am. Chem. Soc.* **117**, 2351 (1995).

94. H. Brennen and co-workers, *J. Bacteriol.* **167**, 66 (1986).

95. J. A. Hoch and co-workers, in A. L. Sonenshein, J. A. Hoch, and R. Losick, eds., *Bacillus subtilis*, American Society for Microbiology, Washington, D.C., 1993, pp. 425.

96. K. G. Maundrell and co-workers, *Yeast* **10**, 1075 (1994).

97. A. Zurlinden and M. E. Schweingruber, *J. Bacteriol.* **176**, 6631 (1994).

98. A. Iwashima and co-workers, *J. Bacteriol.* **172**, 6145 (1990); *ibid.*, **174**, 4701 (1992).

99. A. Schweingruber and co-workers, *Genetics* **130**, 445 (1992).

David Burdick
Hoffmann-La Roche Inc.

VITAMIN A

The curative role of the juice of cooked liver extracts in the treatment of night blindness was first recognized and practiced by the Egyptians in ancient times (1). Several millennia later, in the early part of the twentieth century, many groups identified a lipid-soluble factor found in milk, butter, and egg yolks which was therapeutic for this condition. In 1913, McCollum and Davis coined the term fat-soluble A and subsequently, ascribed the growth-promoting effects of liver extracts to this material (2,3). Later, Osborne and Mendel found that cod liver oil contained an ingredient that was essential for growth promotion in rats (4).

The structure of vitamin A [11103-57-4] and some of the important derivatives are shown in Figure 1. The parent structure is all-*trans*-retinol [68-26-8] and its IUPAC name is (all-*E*)-3,7-dimethyl-9-(2,6,6-trimethyl-1-cyclohexen-1-yl)-2,4,6,8-nonatetraen-1-ol (1). The numbering system for vitamin A derivatives parallels the system used for the carotenoids. In older literature, vitamin A compounds are named as derivatives of trimethyl cyclohexene and the side chain is named as a substituent. For retinoic acid derivatives, the carboxyl group is denoted as C-1 and the trimethyl cyclohexane ring as a substituent on C-9. The structures of vitamin A and β-carotene were elucidated by Karrer in 1930 and several derivatives of the vitamin were prepared by this group (5,6). In 1935, Wald isolated a substance found in the visual pigments of the eye and was able to show that this material was identical with Karrer's retinaldehyde [116-31-4] (5) (7).

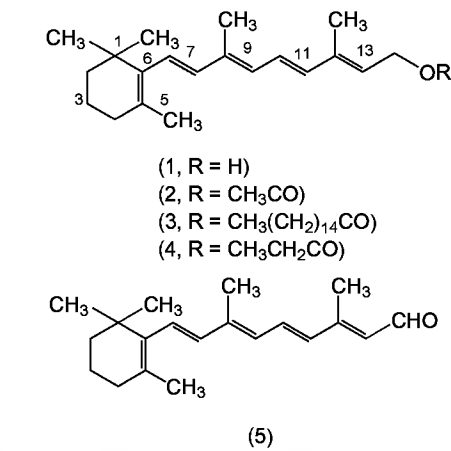
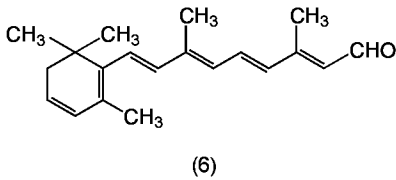


Fig. 1. Vitamin A and derivatives: retinol (1), retinyl acetate [127-47-9] (2), retinyl palmitate [79-81-2] (3), and retinyl propionate [7069-42-3] (4).

Vitamin A₂ [79-80-1] (6) is structurally similar to vitamin A₁ [68-26-8] and is also found in fish oils. This compound is important biologically for fish and other lower animals. Interestingly, tadpoles require vitamin A₂ but after metamorphosis require vitamin A₁ (8).



Vitamin A constitutes the most significant sector of the commercial retinoid market and is used primarily in the feed area. In the pharmaceutical area, there are several important therapeutic dermatologic agents which structurally resemble vitamin A and they are depicted in Figure 2 (see PHARMACEUTICALS). The carotenoids as provitamin A compounds also represent an important commercial class of compounds with β-carotene [7235-40-7] (10) occupying the central role (Fig. 3) (9).

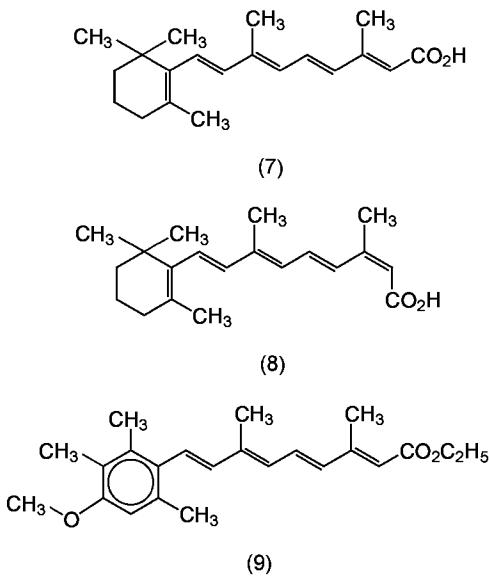


Fig. 2. Commercially important retinoids: retinoic acid [302-79-4] (Tretinoin) (7), 13-Z-retinoic acid [4759-48-2] (Isotretinoin) (8), and etretinate [54350-48-0] (9).

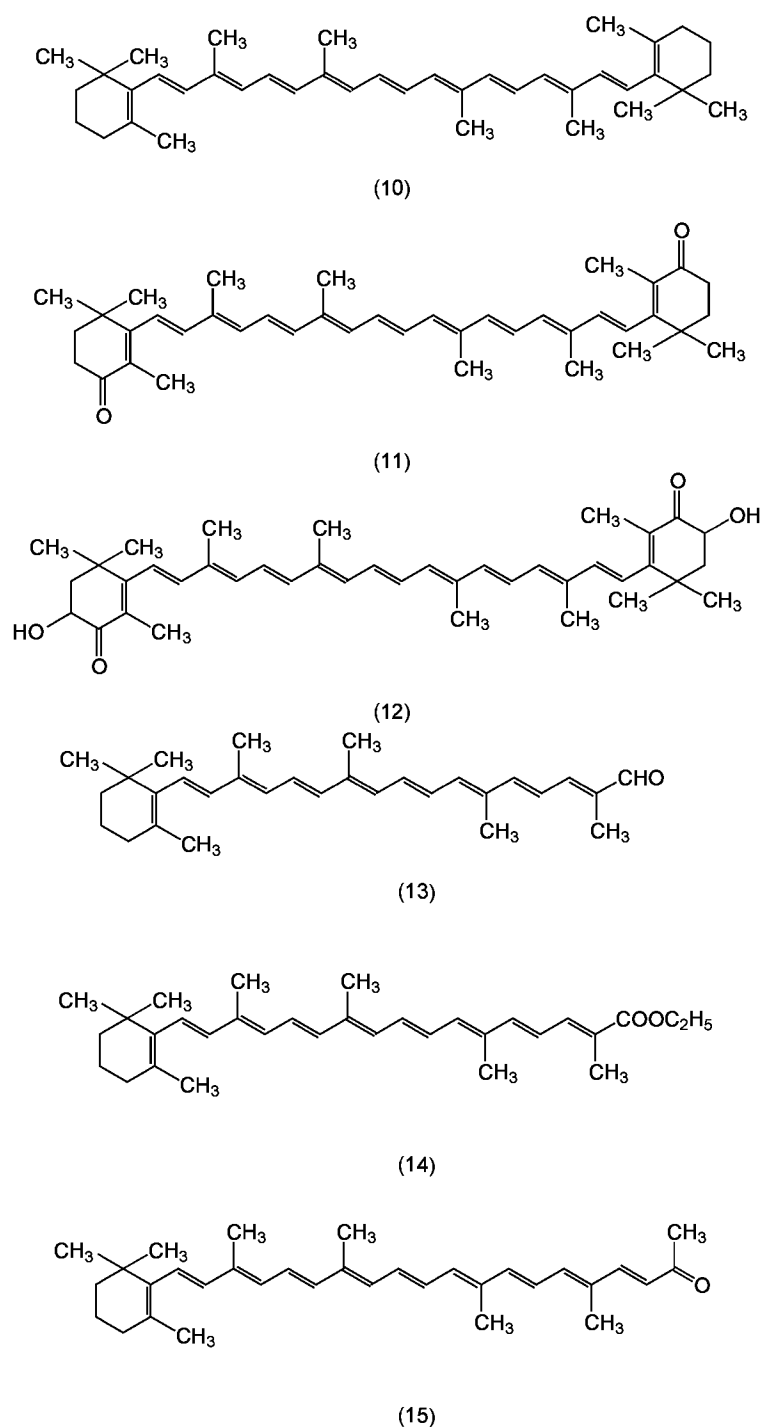
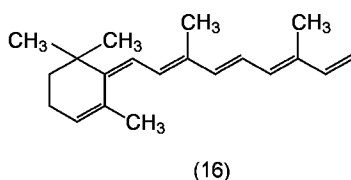


Fig. 3. Commercially important carotenoids: β -carotene [10], canthaxanthin [514-78-3] (11), astaxanthin [472-61-7] (12), β -apo-8'-carotenal [1107-26-2] (13), β -apo-8'-carotenoic acid ethyl ester [1109-11-1] (14), and citranaxanthin [3604-90-8] (15).

Chemical and Physical Properties

Because of the presence of an extended polyene chain, the chemical and physical properties of the retinoids and carotenoids are dominated by this feature. Vitamin A and related substances are yellow compounds which are unstable in the presence of oxygen and light. This decay can be accelerated by heat and trace metals. Retinol is stable to base but is subject to acid-catalyzed dehydration in the presence of dilute acids to yield anhydrovitamin A [1224-18-8] (16). Retro-vitamin A [16729-22-9] (17) is obtained by treatment of retinol in the presence of concentrated hydrobromic acid. In the case of retinoic acid and retinal, reisomerization is possible after conversion to appropriate derivatives such as the acid chloride or the hydroquinone adduct. Table 1 lists the physical properties of β -carotene [7235-40-7] and vitamin A.



(18,19). Their differences lie in methodology to this key intermediate as well as in the methodology to elaborate the side chain. A review of the early work is available (20).

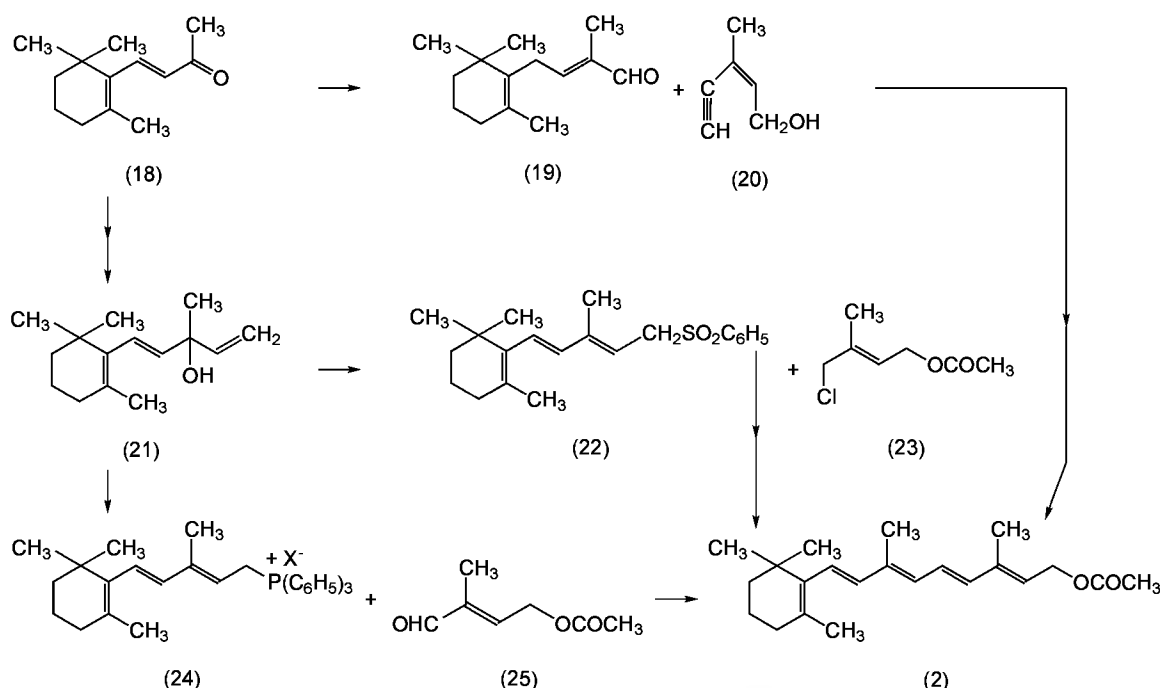


Fig. 4. Commercial synthesis of vitamin A acetate (2).

In the process practiced by Hoffmann-La Roche, a $C_{13} + C_1 + C_6$ strategy is employed. In this approach, β -ionone (18) is subjected to a Darzens' condensation to yield the C_{14} aldehyde [116-31-4] (19) (see Fig. 4). Construction of the side chain is completed by a metal acetylide coupling reaction with compound (20). Acetylation, partial reduction of the triple bond, and acid-catalyzed elimination of water completes the synthesis. BASF and Rh  ne-Poulenc use a different scheme and extend the side chain of β -ionone in an initial step via a Grignard reaction with a metal acetylide. Semihydrogenation yields vinyl β -ionol (21) and it is at this point that the approaches diverge. In the Rh  ne-Poulenc $C_{15} + C_5$ synthesis, the carbon terminus of vinyl β -ionol is activated by conversion to the sulfone (22). In a second step, the anion of the sulfone is reacted with C_5 -chloroacetate (23) to yield vitamin A acetate. BASF utilizes a Wittig olefination and first prepares the phosphonium salt of the C_{15} unit. Reaction of the salt (24) with the C_5 aldehyde (25) leads to vitamin A acetate (21).

Vitamin A palmitate [79-81-2] (3), a commercially important form of the vitamin, is produced from vitamin A acetate (2) via a transesterification reaction with methyl palmitate. Enzymatic preparation of the palmitate from the acetate has also been described (22).

In addition to differences in their methodology to extend the carbon chain, these manufacturers differ in their syntheses of β -ionone. β -Ionone is commercially prepared via an acid-catalyzed rearrangement of pseudoionone (26). This intermediate is manufactured on an industrial scale from either citral (27) or dehydro-linalool (28) (21) (Fig. 5).

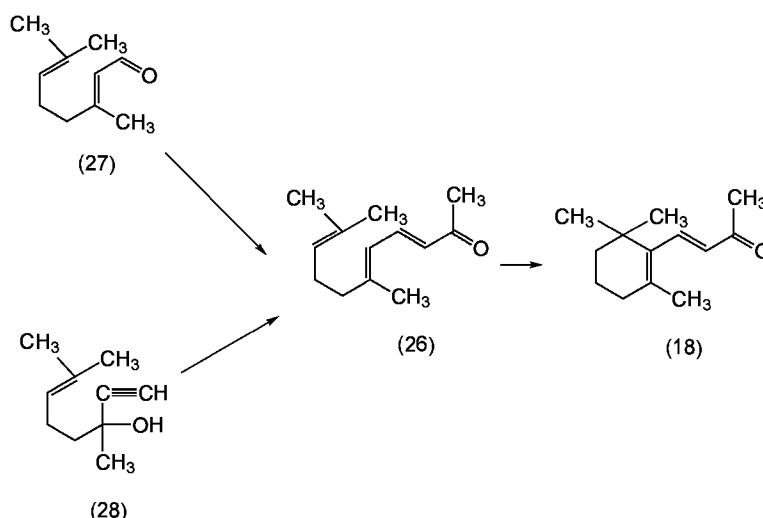


Fig. 5. Industrial approaches to β -ionone (18).

Citral is prepared starting from isobutene and formaldehyde to yield the important C_5 intermediate 3-methylbut-3-enol (29). Pd-catalyzed isomerization affords 3-methylbut-2-enol (30). The second C_5 unit of citral is derived from oxidation of (30) to yield 3-methylbut-2-enal (31). Coupling of these two fragments produces the dienol ether (32) and this is followed by an elegant double Cope rearrangement (21) (Fig. 6).

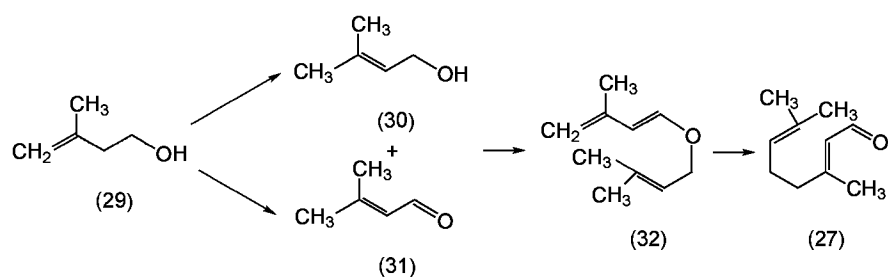
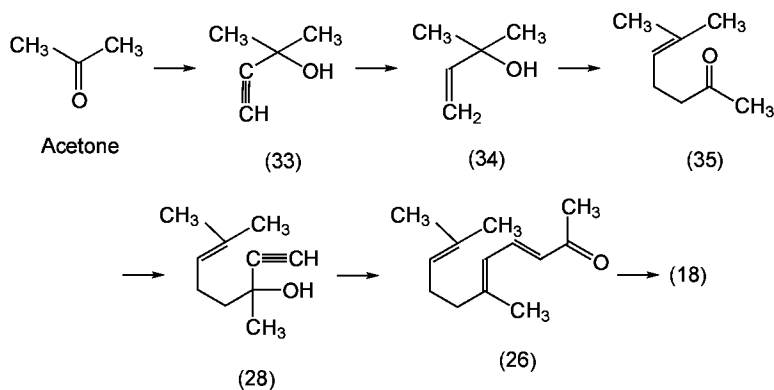


Fig. 6. Synthesis of citral (27).

The synthesis of dehydro-linalool (28) relies on the basic chemicals acetone and acetylene. Addition of a metal acetylide to acetone yields methylbutynol (33). Semihydrogenation affords the alkene (34) which is reacted with *i*-propenylmethyl ether. A Cope rearrangement of the adduct yields methylheptenone (35). Addition of a second mole of metal acetylide to dehydro-linalool (28) is followed by a second Cope rearrangement to yield pseudoionone (26) (9,21) (Fig. 7).

Fig. 7. Acetone-based synthesis of β -ionone.

In other work, sulfone chemistry plays an integral part of the syntheses of both β -carotene and vitamin A by workers at Kuraray. In this approach, the anion of C_{10} β -cyclogeranyl sulfone (36) is condensed with the C_{10} aldehyde (37). The resulting β -hydroxy sulfone (38) is treated with dihydropyran followed by a double elimination to yield vitamin A acetate. Alternatively, the β -hydroxy sulfone (38) can be converted to the δ -halo sulfone (39) and a similar double elimination scheme is employed (23,24) (Fig. 8).

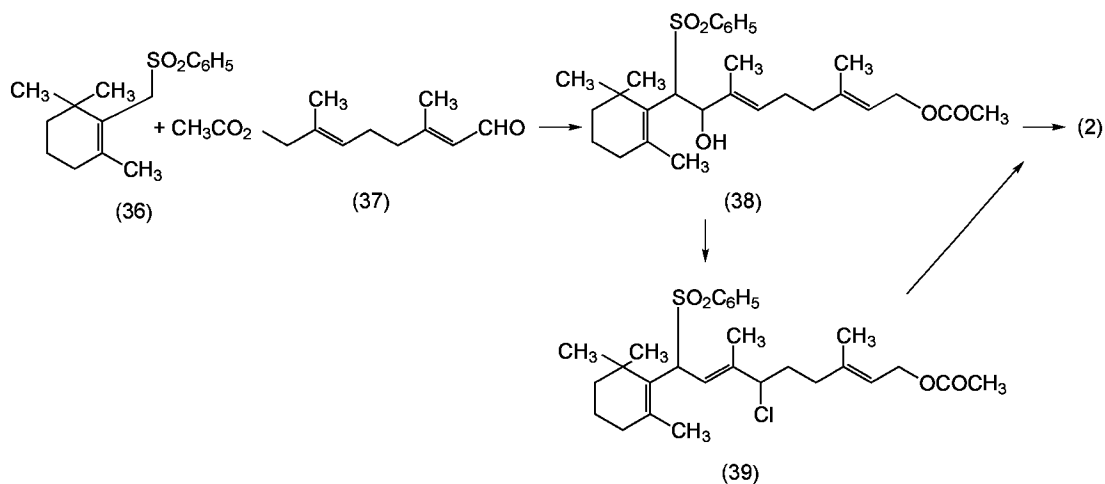


Fig. 8. Kuraray synthesis.

Work at Rhône-Poulenc has involved a different approach to retinal and is based on the palladium-catalyzed rearrangement of the mixed carbonate (41) to the allenyl enal (42). Isomerization of the allene (42) to the polyene (43) completes the construction of the carbon framework. Acid-catalyzed isomerization yields retinal (5). A decided advantage of this route is that no by-products such as triphenylphosphine oxide or sodium phenylsulfinate are formed. However, significant yield improvements would be necessary for this process to compete with the current commercial syntheses (25–27) (Fig. 9).

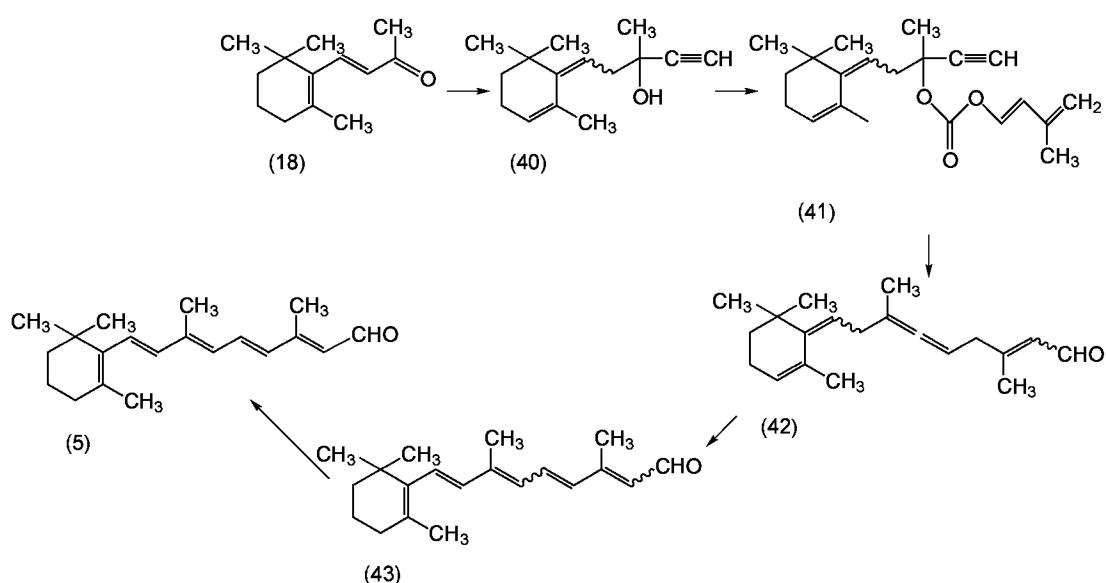
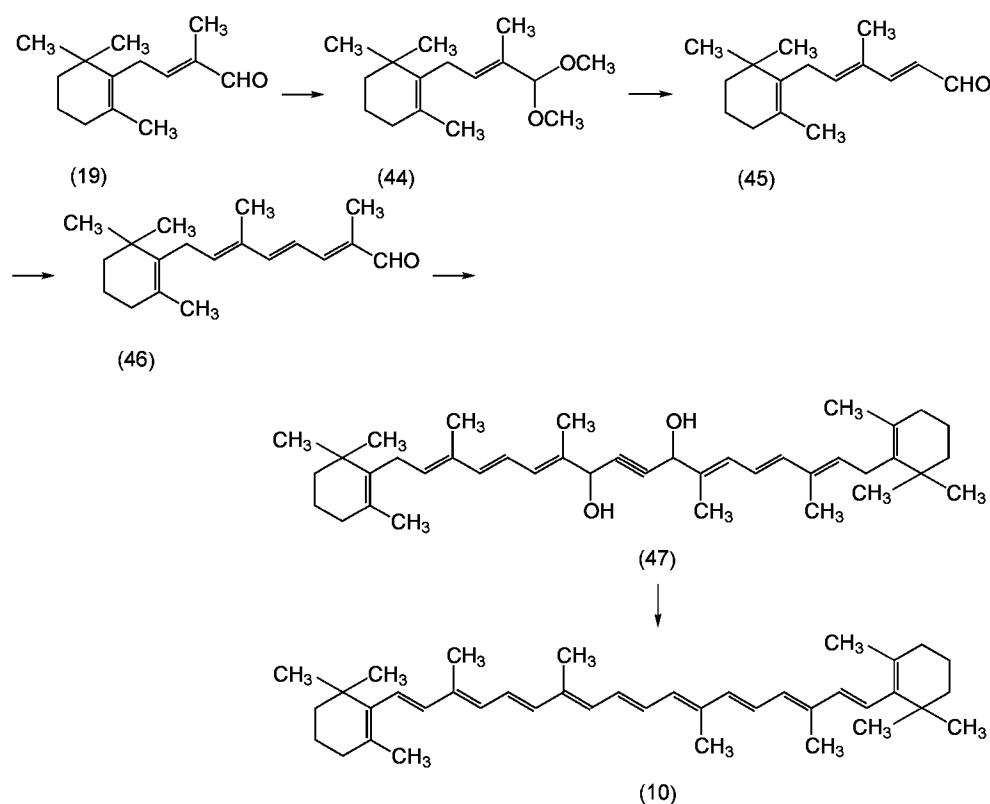


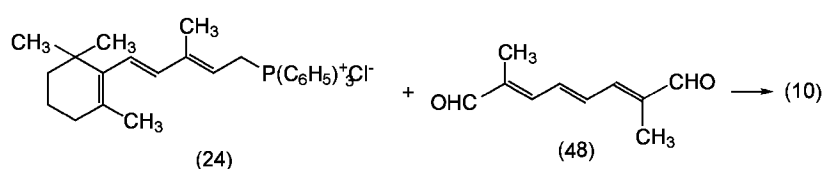
Fig. 9. Allene synthesis of vitamin A aldehyde.

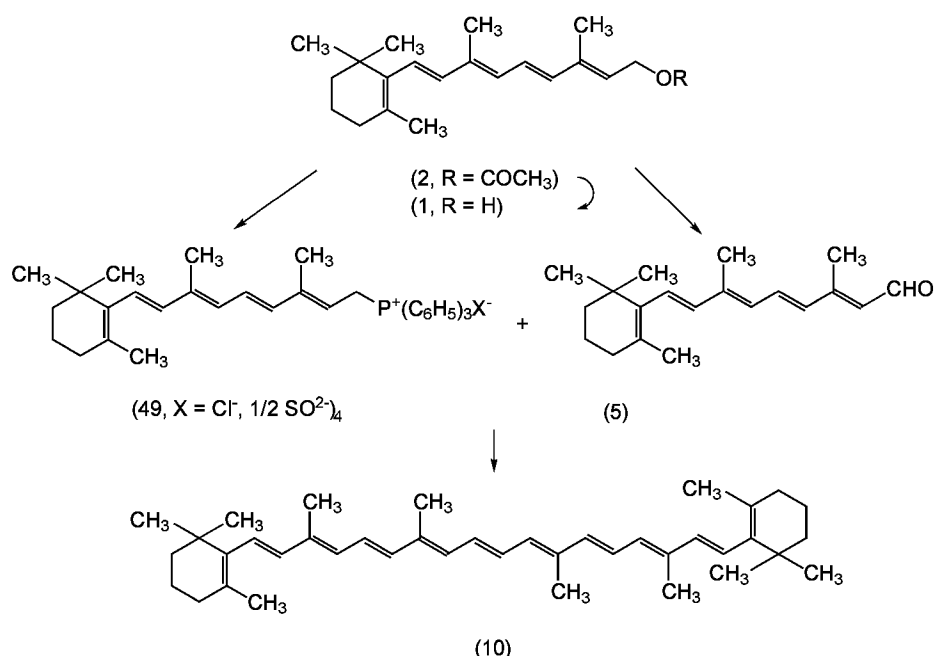
In contrast to the similarities seen in the majority of the industrial syntheses of vitamin A, substantial diversity is observed in the preparations of the carotenoids. Owing to the fact that all-*trans* stereochemistry is required in the final product and that many olefination reactions yield the *cis* product, the ease of isomerization at a given locus on the polyene chain has influenced the choice of building blocks. In addition, technological strengths in the construction of these building blocks as well as synergies with a core olefination strength have played an important role in the choice of synthesis.

Hoffmann-La Roche has produced β -carotene since the 1950s and has relied on core knowledge of vitamin A chemistry for the synthesis of this target. In this approach, a five-carbon homologation of C_{14} vitamin A aldehyde (19) is accomplished by successive acetalizations and enol ether condensations to prepare the C_{10} aldehyde (46). Metal acetylide coupling with two molecules of aldehyde (46) completes construction of the C_{40} carbon framework. Selective reduction of the internal triple bond of (47) is followed by dehydration and thermal isomerization to yield β -carotene (21) (Fig. 10).

Fig. 10. Hoffmann-La Roche synthesis of β -carotene.

In the BASF synthesis, a Wittig reaction between two moles of C_{15} phosphonium salt (vitamin A intermediate (24)) and C_{10} dialdehyde (48) is the important synthetic step (9,28,29). Thermal isomerization affords all *trans*- β -carotene (Fig. 11). In an alternative preparation by Roche, vitamin A process streams can be used and in this scheme, retinol is carefully oxidized to retinal, and a second portion is converted to the C_{20} phosphonium salt (49). These two halves are united using standard Wittig chemistry (8) (Fig. 12).

Fig. 11. BASF synthesis of β -carotene.

Fig. 12. Alternative Roche synthesis of β -carotene.

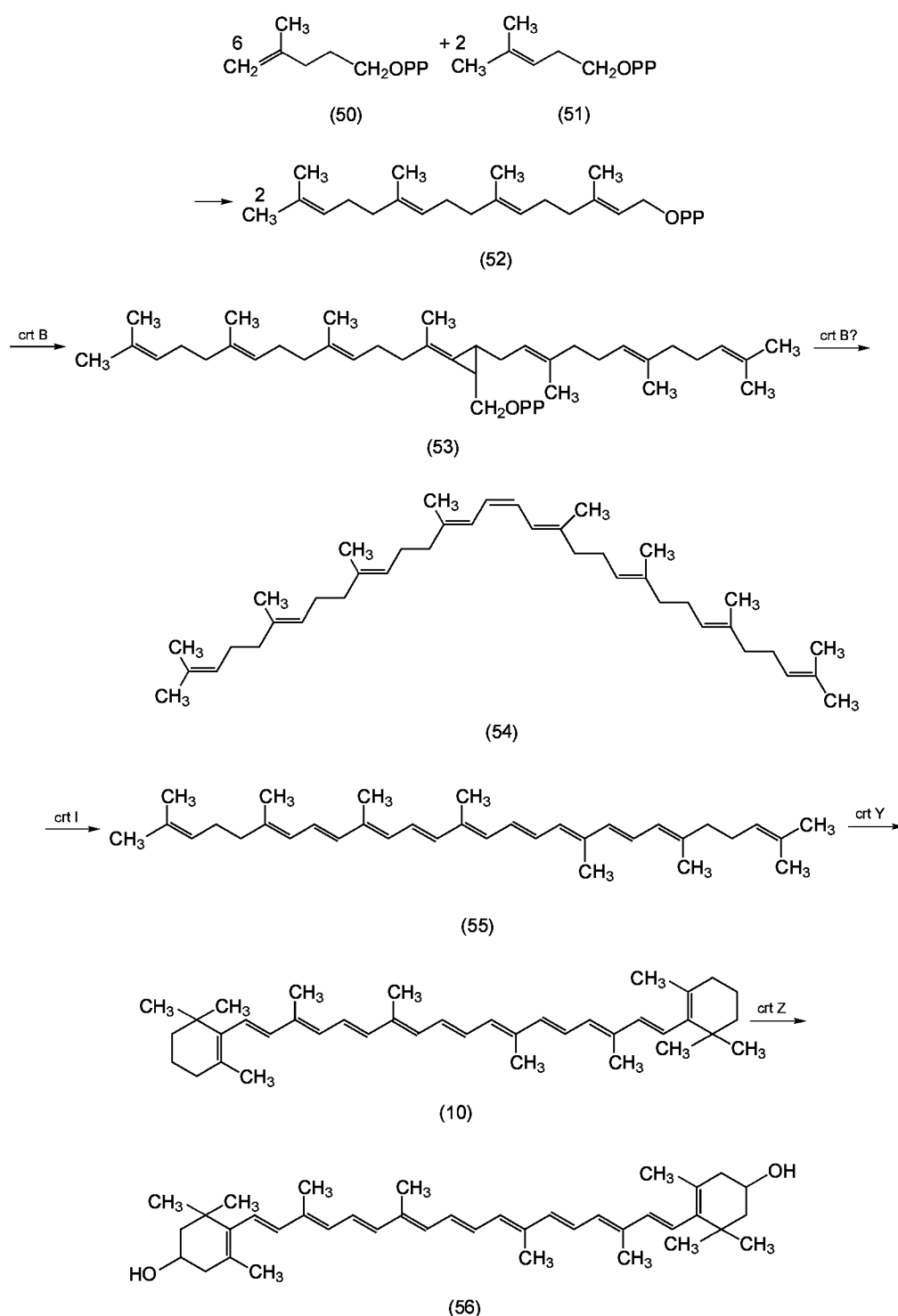
Other approaches to direct C_{20} couplings have been reported (9,30–35). Based on their knowledge of sulfone chemistry, Rhône-Poulenc has patented many syntheses of β -carotene which use this olefination chemistry (36–41). Horner-Emmons chemistry has also been employed for this purpose (42). The synthetic approaches to the carotenoids have been reviewed (43).

Compounds labeled ^{14}C , ^{13}C , ^3H , or ^2H are extremely important in understanding the absorption, distribution, metabolism, and excretion of these materials in biological systems. The preparation of these materials has been reviewed (44,45) and has been the subject of recent investigations (46).

Biosynthesis

In nature, vitamin A aldehyde is produced by the oxidative cleavage of β -carotene by 15,15'- β -carotene dioxygenase. Alternatively, retinal is produced by oxidative cleavage of β -carotene to β -apo-8'-carotenal followed by cleavage at the 15,15'-double bond to vitamin A aldehyde (47). Carotenoid biosynthesis and fermentation have been extensively studied both in academic as well as in industrial laboratories. On the commercial side, the focus of these investigations has been to increase fermentation titers by both classical and recombinant means.

The carotenoid skeleton is assembled in a primary step by the coupling of two moles of geranylgeranyl pyrophosphate (52) by an enzyme that is encoded by the *crt B* (carotenogenic) gene. The resulting prephytoene pyrophosphate (53) is further transformed to phytoene (54) possibly by products also derived from the *crt B* gene. Phytoene is a common intermediate in carotenoid biosynthesis. For example, in *E. uredozona* phytoene is converted to lycopene (55) in sequential dehydration steps. These reactions are catalyzed by an enzyme called phytoene desaturase, which is encoded on the gene cluster *crt I*. Interestingly, only *cis*-phytoene and *trans*-lycopene are detected, and it is hypothesized that *trans*-phytoene is formed in a nonenzymatic step. By further biosynthetic transformations, β -carotene (10) is produced from lycopene and zeaxanthin (56) from β -carotene (Fig. 13) (48).

Fig. 13. Biosynthetic transformation of β -carotene.

The majority of industrial research describes classical approaches to yield improvement (49). However, there has been some work using genetically modified organisms. In the case of these recombinant organisms, the carotenoid biosynthetic gene cluster has been expressed in noncarotogenic species such as *E. coli* (50) and *S. cerevisiae* (51).

Analytical Methods

Biological, spectroscopic, and chromatographic methods have been used to assay vitamin A and the carotenoids. Biological methods have traditionally been based on the growth response of vitamin A-deficient rats. The utility and shortcomings of this test have been reviewed (52,53). This test has found applicability for analogues of retinol (54,55). Carotenoids that function as provitamin A precursors can also be assayed by this test (56).

Spectroscopic methods such as uv and fluorescence have relied on the polyene chromophore of vitamin A as a basis for analysis. Indirectly, the classical Carr-Price colorimetric test also exploits this feature and measures the amount of a transient blue complex at 620 nm which is formed when vitamin A is dehydrated in the presence of Lewis acids. For uv measurements of retinol, retinyl acetate, and retinyl palmitate, analysis is done at 325 nm. More sensitive measurements can be obtained by fluorescence. Excitation is done at 325 nm and emission at 470 nm. Although useful, all of these methods suffer from the fact that the method is not specific and any compound which has spectral characteristics similar to vitamin A will assay like the vitamin (57).

More specific methods involve chromatographic separation of the retinoids and carotenoids followed by an appropriate detection method. This subject has been reviewed (57). Typically, hplc techniques are used and are coupled with detection by uv. For the retinoids, fluorescent detection is possible and picogram quantities of retinol in plasma have been measured (58–62). These techniques are particularly powerful for the separation of isomers. Owing to the thermal lability of these compounds, gc methods have also been used but to a lesser extent. Recently, the utility of cool-on-column injection methods for these materials has been demonstrated (63).

Owing to the light and air sensitivity of the carotenoids and retinoids, sample handling is a critical issue. It is recommended to conduct extraction of these materials with peroxide-free solvents, to store biological samples at -70°C under argon and in the dark, to perform the analysis under yellow light, and to use reference compounds of high purity (57).

In the *United States Pharmacopeia* (USP), vitamin A content is determined by the ratio of the corrected absorbance to the observed absorbance and is

not to be less than 0.85 at 325 nm. Total vitamin A content is to be 95° of label claim. For β-carotene, the assay is performed using uv spectroscopy and is determined by the absorbance at 455 nm. The range of the assay should be from 96.0 to 101.0° (64).

Occurrence

Rich sources of vitamin A include dairy products such as milk, cheese, butter, and ice cream. Eggs as well as internal organs such as the liver, kidney, and heart also represent good sources. In addition, fish such as herring, sardines, and tuna, and in particular the liver oil from certain marine organisms, are excellent sources. Because the vitamin A in these food products is derived from dietary carotenoids, vitamin A content can vary considerably. Variation of vitamin A content in food can also result from food processing and in particular, oxidation processes (8).

Fertile sources of carotenoids include carrots and leafy green vegetables such as spinach. Tomatoes contain significant amounts of the red carotenoid, lycopene. Although lycopene has no vitamin A activity, it is a particularly efficient antioxidant (see ANTIOXIDANTS). Oxidation of carotenoids to biologically inactive xanthophylls represents an important degradation pathway for these compounds (56).

Biological Functions

In humans, vitamin A has important functions in vision, growth, and tissue differentiation. Its role in vision is fairly well understood and involves initial absorption of all-*trans*-retinol by the pigment epithelium cells of the eye. This event is followed by isomerization to 11-*cis*-retinol by all-*trans*-11-*cis*-retinol isomerase. After conversion of 11-*cis*-retinol to 11-*cis*-retinal, reaction with opsin forms rhodopsin. Absorption of light by this pigment generates the visual signal by *cis* trans isomerization. After release of *trans*-retinal from the pigment, *trans*-retinal is rapidly reduced to all-*trans*-retinol and the cycle repeats itself. Consequently, vitamin A is not consumed in the process. One feature of this system is that the basic chemistry is common to many photosensitive systems from such diverse species as halophilic bacteria to higher organisms such as vertebrates (65).

As previously described, the biological assay for vitamin A was based on the growth response of rats fed a vitamin A-deficient diet. Vitamin A is necessary for normal bone growth and remodeling and is required for the activity of epiphyseal cartilage. Retinoic acid is the active form of vitamin A for this function and it has been observed that ingestion of retinoic acid restores growth but cannot maintain vision (8). In animals cycled on retinoic acid, ie, fed a retinoic acid-containing diet then deprived of a retinoic acid-containing diet, these animals become sensitive to the removal of retinoic acid. This condition is characterized by a loss in appetite followed by a depression in growth. As a result, loss of appetite is one of the first symptoms of a vitamin A-deficient animal (8).

The specific role of vitamin A in tissue differentiation has been an active area of research. The current thinking, developed in 1979, involves initial delivery of retinol by *holo*-RBP (retinol-binding protein) to the cell cytosol (66). Retinol is then ultimately oxidized to retinoic acid and binds to a specific cellular retinoid-binding protein and is transported to the nucleus. Retinoic acid is then transferred to a nuclear retinoic acid receptor (RAR), which enhances the expression of a specific region of the genome. Transcription occurs and new proteins appear during the retinoic acid-induced differentiation of cells (56).

In contrast to vitamin A, the carotenoids are important in both plant and animal kingdoms. In plants, carotenoids are associated with photosynthetic structures. The function of these pigments is twofold: (1) the pigments serve as acceptors and act as energy-transfer agents and allow for excitation energy to be transferred from the carotenoid to the porphyrin; (2) the pigments protect the organism from light-induced photooxidative damage. In this regard, protection is achieved by two mechanisms: quenching of the excited triplet state of chlorophyll (itself a producer of singlet oxygen), and quenching of singlet oxygen. In both cases, a triplet excited state of the carotenoid, which decays back to the ground state by a radiationless conversion, is produced (67).

Many carotenoids function in humans as vitamin A precursors; however, not all carotenoids have provitamin A activity (Table 3). Of the biologically active carotenoids, β-carotene has the greatest activity. Despite the fact that theoretically one molecule of β-carotene is a biological source of two molecules of vitamin A, this relationship is not observed and 6 μg β-carotene is equivalent to 1 μ vitamin A. Although β-carotene and vitamin A have complementary activities, they cannot totally replace each other. Because the conversion of β-carotene to vitamin A is highly regulated, toxic quantities of vitamin A cannot accumulate and β-carotene can be considered as a safe form of vitamin A (8).

Table 3. Provitamin A Activity of Selected Carotenoids^a

Provitamin A activity	No provitamin A activity
β-carotene	astaxanthin
α-carotene	canthaxanthin
γ-carotene	lutein
β-cryptoxanthin	lycopene
β-zeacarotene	zeaxanthin

^a Ref. 8.

Owing to the presence of an extended polyene chain, all carotenoids are effective single-oxygen quenchers and antioxidants. In addition, they can stimulate the immune response and, as a result, may protect against certain forms of cancer. These separate functions, ie, vitamin A equivalents or antioxidants, have allowed for carotenoids to be classified according to whether they are nutritionally or biologically active. For example, β-carotene is both nutritionally and biologically active whereas 14'-β-apocarotenal is nutritionally active but biologically inactive. Lycopene is nutritionally inactive but biologically active, whereas phytoene, the natural precursor to β-carotene, is both nutritionally and biologically inactive (56).

Requirements

Animals cannot synthesize vitamin A-active compounds and necessary quantities are obtained by ingestion of vitamin A or by consumption of appropriate provitamin A compounds such as β-carotene. Carotenoids are manufactured exclusively by plants and photosynthetic bacteria. Until the discovery of vitamin A in the purple bacterium *Halobacterium halobium* in the 1970s, vitamin A was thought to be confined to only the animal kingdom (56). Table 4 lists RDA and U.S. RDA for vitamin A (67).

Table 4. Recommended Daily Allowances of Vitamin A^a

Sex age group	1989 RDA (RE) ^b	U.S. RDA (IU) ^c
infants <12 months	375	1500
children		
1–3 yr	400	2500
4–6 yr	500	5000
7–10 yr	700	5000
males >11 yr	1000	5000

females >11 yr	800	5000
additional during pregnancy		+3000
additional during lactation		+3000

^a Ref. 8.
^b RE retinol equivalents (1 RE = 1 µg of all-*trans*-retinol).
^c IU international units (1 IU = 0.30 µg of all-*trans*-retinol).

Deficiency

In humans, vitamin A deficiency manifests itself in the following ways: night blindness, xerophthalmia, Bitot's spots, and corneal involvement and ulceration. Changes in the skin have also been observed. Although vitamin A deficiency is seen in adults, the condition is particularly harmful in the very young. Often, this results from malnutrition (56).
On a vitamin A-deficient diet, mucus-secreting tissues become keratinized. This condition tends to occur in the trachea, the skin, the salivary glands, the cornea, and the testes. When this occurs in the cornea, it can be followed by blindness. Vitamin A deficiency is the principal cause of blindness in the very young. This problem is particularly acute in the third world (8).

Safety

Vitamin A toxicity can be categorized as either acute or chronic. Acute toxicity results from extremely high doses (>500,000 IU of vitamin A). In children, approximately half of that amount causes problems. Hypervitaminosis A is characterized by headache, blurred vision, loss of coordination, nausea, and peeling and itchy skin. Chronic vitamin A toxicity occurs in adults with long-term intakes of >50,000 IU d. Symptoms include dry hair, hair loss, weakness, headache, bone thickening, enlarged liver and spleen, anemia, abnormal menstrual periods, stiffness, and joint pain. Most of these symptoms are reversible. In animals, extremely high doses of vitamin are teratogenic (56). On the other hand, the carotenoids are generally nontoxic and there have been only a few isolated cases of problems associated with a large daily intake. Tanning pills that contain large doses of canthaxanthin were shown to cause canthaxanthin retinopathy in patients with erythropoietic porphyria (68,70).

Uses

Vitamin A is generally used in feeds, foods, and pharmaceutical applications; however, the principal use of carotenoids is as a colorant. β-Carotene, for example, is used to color fat products such as margarine, shortening, butter, cheese, egg nog, and ice cream. It is also used in baked goods, pasta products, juices, and beverages. It imparts a natural yellow to orange shade. Conversely, if an orange to reddish orange color is desired, the carotenoid of choice is β-apo-8'-carotenal [1107-26-2]. This carotenoid is used to color juices, fruit drinks, soups, jams, jellies, and gelatin (qv). Carotenoids are also added to feed to color poultry, fish, and of egg yolks. In order to guard against oxidative damage during feed processing, these compounds are protected in a gelatin matrix as beadlets (see COLORANTS FOR FOOD, DRUGS, COSMETICS, AND MEDICAL DEVICES).

Economic Aspects

Vitamin A is manufactured by Hoffmann-La Roche (Switzerland), BASF (Germany), and Rhône-Poulenc (France), as well as by some smaller suppliers in India, China, and Russia. The worldwide production is estimated to be 2500 to 3000 metric tons. About three-quarters of this production is for animal feed; the remainder is for food fortification and pharmaceuticals (qv). The main trade names of feed products are Rovimix, Lutavit, and Microvit. Prices depend on application forms and are approximately \$60–\$70 10⁹ IU retinol (1995); ie, \$200–\$233 10⁹ RE. One IU is equivalent to 0.300 µg of all-*trans*-retinol and 1 RE is equivalent to 1 µg of all-*trans*-retinol.

BIBLIOGRAPHY

"Vitamin A" under "Vitamins" in *ECT* 1st ed., Vol. 14, pp. 791–813, by W. Oroshnik, Ortho Research Foundation; in *ECT* 2nd ed., Vol. 21, pp. 490–509, by W. Oroshnik, Chemo Dynamics; in *ECT* 3rd ed., Vol. 24, pp. 140–158, by O. Isler and F. Kienzle, F. Hoffmann LaRoche & Co.

- G. Wolf, *Am. J. Clin. Nutr.* **31**, 290 (1978).
- E. V. McCollum and M. Davis, *J. Biol. Chem.* **15**, 167 (1913).
- E. V. McCollum and M. Davis, *J. Biol. Chem.* **23**, 291 (1915).
- T. B. Osborne and L. B. Mendel, *J. Biol Chem.* **37**, 187 (1923).
- P. Karrer, A. Helfensten, H. Wehdi, and A. Wettstein, *Helv. Chim. Acta* **13**, 1084 (1930).
- P. Karrer, R. Morf, and K. Schopp, *Helv. Chim. Acta* **14**, 1431 (1931).
- G. Wald, *Nature (London)* **134**, 65 (1934).
- D. Bhatia, ed., in *Encyclopedia of Food Science and Technology*, John Wiley & Sons, Inc., New York, 1991, pp. 2701–2712.
- J. Paust, *Pure Appl. Chem.* **63**, 45 (1991).
- Technical data, Hoffmann-La Roche, Inc., Nutley, N.J., 1996.
- R. A. Heyman, D. J. Mangelsdorf, J. A. Dyck, R. B. Stein, G. Eichelle, R. M. Evans, and C. Thaller, *Cell* **68**, 397 (1992).
- R. S. H. Liu and A. E. Asato, *Tetrahedron*, **40**, 1931 (1984); Y. Bennani, *J. Org. Chem.* **61**, 3542 (1996); A. E. Asato, A. Kini, M. Denney, and R. S. H. Liu, *J. Am. Chem. Soc.* **105**, 2923 (1983); G. Wald, P. K. Brown, R. Hubbard, and W. Oroshnik, *Natl. Acad. Sci. USA*, **41**, 438 (1955); G. Wald, P. K. Brown, R. Hubbard, and W. Oroshnik, *Natl. Acad. Sci. USA*, **42**, 578 (1956); K. Ramamurthy and R. Liu, *Tetrahedron*, **31**, 201 (1975); A. E. Asato and R. S. H. Liu, *J. Am. Chem. Soc.* **97**, 4128 (1975); R. S. H. Liu and A. E. Asato, *Methods. Enzymol.* **88**, 506 (1982); A. Kini, H. Matsumoto, and R. S. H. Liu, *J. Am. Chem. Soc.* **107**, 5078 (1979); A. Kini, H. Matsumoto, and R. S. H. Liu, *Biorg. Chem.* **9**, 406 (1980); G. G. Ladson, S. C. Cary, and W. H. Okamura, *J. Am. Chem. Soc.* **102**, 6355 (1980).
- G. Britton, in G. Britton, S. Liaaen-Jensen, and H. Pfander, eds., *Carotenoids, Vol. 1B: Spectroscopy*, Birkhauseer, Basel, Switzerland, 1995, pp. 13–62.
- G. Britton, G. Armit, S. Y. M. Lau, A. K. Patel, and C. S. Shone, in G. Britton and T. W. Goodwin, eds., *Carotenoid Chemistry and Biochemistry*, Pergamon Press, Oxford, U.K., 1984, pp. 237–251.
- P. Zagalsky, in Ref. 13, *Vol. 1A: Isolation and Analysis*, pp. 287–294.
- S. Liaaen-Jensen, in Ref. 13, pp. 343–354.
- C. H. Eugster, in Ref. 15, pp. 343–354.
- O. Isler, W. Huber, A. Ronoco, and M. Kofler, *Helv. Chim. Acta*, **30**, 1911 (1942).
- O. Isler, A. Ronco, W. Guex, N. C. Hindley, W. Huber, K. Dialer, and M. Kofler, *Helv. Chim. Acta*, **32**, 489 (1949).
- O. Isler, *Pure Appl. Chem.* **51**, 447 (1979).
- C. Mercier and P. Chabardes, *Pure Appl. Chem.* **66**, 1509 (1994).
- Jpn. Pat. 86-93840** (1986), Y. Inada (to Bihaman, Hishara).
- J. Otera, H. Misana, T. Mandai, T. Onishi, S. Suzuki, and Y. Fujita, *Chem. Lett.*, 1883 (1985).

24. J. Otera, H. Misawa, T. Onishi, S. Suzuki, and Y. Fujita, *J. Org. Chem.* **51**, 3834 (1986).

25. H. Bienayme, *Tetrahedron Lett.* **35**, 7387 (1994).

26. *Ibid.*, p. 7383.

27. H. Bienayme, *Bull. Soc. Chim. Fr.* **132**, 696 (1995).

28. H. Pommer and P. C. Thieme, *Top. Curr. Chem.* **109**, 165 (1983).

29. A. Nurrenbach, J. Paust, H. Pommer, J. Scheider, and B. Schultz, *Liebigs Ann. Chem.*, 1146 (1972).

30. H. J. Bestman, C. Kisielowski, and N. Distler, *Angew. Chem.* **88**, 297 (1976).

31. **Ger. Pat. 2,551,914** (1975), H. J. Bestmann and N. Distler (to Hoechst AG).

32. **Ger. Pat. 2,702,633** (1977), B. Schulz, J. Paust, and J. M. Schneider (to BASF AG).

33. **Jpn. Pat. 76-108106** (1979), M. Mukaiyama and A. Ishida (to Japan); A. Ishida, T. Mukaiyama, *Chem. Lett.*, 1127–1130 (1972).

34. **Ger. Pat. 2,641,075** (1977), J. McMurray (to University of California).

35. **Ger. Pat. 2,515,011** (1976), J. McMurray (to University of California).

36. **Ger. Pat. 2,224,606** (1971), P. Chabardes and J. Marc (to Rhône-Poulenc).

37. **Ger. Pat. 2,264,501** (1971), P. Chabardes and M. Julia (to Rhône-Poulenc).

38. **Ger. Pat. 2,305,217** (1972), P. Chabardes, M. Julia, and A. Menet (to Rhône-Poulenc).

39. **Ger. Pat. 2,355,898** (1972), P. Chabardes, M. Julia, and A. Menet (to Rhône-Poulenc).

40. **Ger. Pat. 2,305,267** (1972), P. Chabardes, M. Julia, and A. Menet (to Rhône-Poulenc).

41. **Ger. Pat. 2,202,689** (1971), M. Julia (to Rhône-Poulenc).

42. **U.S. Pat. 4,916,250** (Apr. 10, 1990), J. H. Babler (to University of Loyola); J. H. Babler and S. A. Schlidt, *Tetrahedron Lett.* **33**, 7697–7700 (1992).

43. K. Bernhard and H. Mayer, *Pure Appl. Chem.* **63**, 35–44 (1991); E. Widmer, *Pure Appl. Chem.* **57**, 741–752 (1985).

44. O. Isler, R. Røegg, U. Schwieter, and J. Wursch, *Vitam. Horm.* (NY), **18**, 295 (1960).

45. H. H. Kaegi, in M. B. Sporn, A. R. Roberts, and D. S. Goodman, eds., *The Retinoids*, Vol. I, Academic Press, Orlando, Fla., 1984, pp. 147–174.

46. M. F. Boehm and co-workers, *J. Med. Chem.* **37**, 408–414 (1994); I. M. Mishina, E. N. Kamaukhova, V. I. Shvets, and M. G. Vladimirova, *Biokhnologiya*, 13–17 (1993); A. A. Liebman, W. Burger, S. C. Choudhry, and J. Cupano, *Methods Enzymol.* **213**(A), 42–49 (1992); M. Groesbeek and J. Lugtenberg, *Photochem. Photobiol.* **56**, 903–908 (1992); R. W. Curley, Jr., M. A. Blust, and K. A. Humphries, *J. Labelled Compd. Radiopharm.* **29**, 1331–1335 (1991); E. M. M. Van den Berg, A. Van der Bent, and J. Lugtenburg, *Recl. Trav. Chim. Pays-Bas* **109**, 160–167 (1990); H. Akita, *Chem. Pharm. Bull.* **31**, 1796–1799 (1983); S. C. Welch and J. M. Gruber, *J. Org. Chem.* **47**, 385–389 (1982).

47. J. Ganguly, *Biochemistry of Vitamin A*, CRC Press, Boca Raton, Fla., 1989.

48. N. Misawa, M. Nakagawa, K. Kobayashi, S. Yamano, Y. Izawa, K. Nakamura, and K. Harahima, *J. Bacteriol.* **172**, 6704 (1990).

49. **World Pat. 9406918** (1994), A. J. van Ooyen (to Gist-Brocades, Inc.); **Eur. Pat. 635576 A1 950125** (1995), T. Kiyota, M. Takaki, A. Tsubokura, and H. Yoneda (to Nippon Oil KK); **Eur. Pat. 454024 A 911030** (1991), W. D. Prevatt and R. E. Torregrossa (to Phillips Petroleum Co.); **World Pat. 9103571** (1991), D. L. Gierhart (to Applied Food BioTechnology, Inc.).

50. **Eur. Pat. 107493 900420** (1990), N. Misawa, K. Kobayashi, and K. Nakamura (to Kirin Brewery Co, Ltd.).

51. **World Pat. 9113078 A1 91095** (1991), R. L. Ausich, F. L. Brinkhaus, I. Mukharji, J. H. Proffitt, J. G. Yarger, and B. Yen (to Amoco Corp.).

52. N. D. Embree, S. R. Ames, R. W. Lehman, and P. L. Harris, *Meth. Biochem. Anal.* **4**, 43 (1957).

53. P. L. Harris, *Vitam. Horm.* **18**, 341 (1960).

54. P. Tosukhowong and J. A. Olson, *Biochem. Biophys. Acta*, **529**, 438 (1978).

55. A. B. Barua and J. A. Olson, *Biochem. J.* **244**, 231 (1987).

56. J. A. Olson, in L. J. Machlin, ed., *The Handbook of Vitamins*, Marcel Dekker, New York, 1991, pp. 1–57.

57. H. C. Furr, A. B. Barua, and J. A. Olson, in A. P. De Leenheer, W. F. Lambert, and H. J. Nelis, eds., *Modern Chromatographic Analysis of the Vitamins*, Marcel Dekker, New York, 1992, pp. 1–71.

58. R. Schindler, A. Kloop, C. Gorny, and W. Feldheim, *Int. J. Vit. Nutr. Res.* **55**, 25 (1985).

59. A. T. Rhys Williams, *J. Chromatogr.* **341**, 198 (1985).

60. A. J. Speek, C. Wongkham, N. Limratana, W. Saowakontha, and W. H. Schreurs, *Curr. Eye. Res.* **5**, 841 (1986).

61. C. A. Collins and C. K. Chow, *J. Chromatogr.* **317**, 349 (1984).

62. H. K. Biesalski and H. Weiser, *J. Clin. Chem. Clin. Biochem.* **27**, 65 (1989).

63. C. R. Smidt, A. D. Jones, and A. J. Clifford, *J. Chromatogr.* **21**, 434 (1988).

64. *The United States Pharmacopeia XXII* (USP XXII-NF XVII), United States Pharmacopeial Convention, Inc., Rockville, Md., 1990, pp. 156–157, 1451, 1550–1551.

65. C. D. B. Bridges, in Ref. 45, pp. 126–176.

66. F. Chytil and D. Ong, *Fed. Proc.* **38**, 2510 (1979).

67. P. Mathis and C. C. Schenck, in Ref. 14, pp. 339–351.

68. Committee on Dietary Allowances, Food and Nutrition Board, *Recommended Dietary Allowances*, 10th ed., National Academy Press, Washington, D.C., 1989.

69. U. Weber, G. Goerz, and R. Hennekes, *Klin. Monatsbl. Augenheilkd.* **186**, 351 (1985).

70. B. Daicker, K. Schiedt, J. J. Adnet, and P. Bermond, *Graefe's Arch. Clin. Exp. Ophthalmol.* **225**, 189 (1987).

General References

D. Bhatia, ed., "Vitamins, Part III: Vitamin A and Provitamin Carotenoids," in the *Encyclopedia of Food Science and Technology*, John Wiley & Sons, Inc., New York, 1991.

J. Ganguly, *Biochemistry of Vitamin A*, CRC Press, Boca Raton, Fla., 1989.

A. P. De Leenheer, W. F. Lambert, and H. J. Nelis, eds., *Retinoids and Carotenoids in Modern Chromatographic Analysis of the Vitamins*, Marcel Dekker, New York, 1992.

M. A. Livrea and G. Vidali, eds., *Retinoids: From Basic Science to Clinical Applications*, Birkhauser Verlag, Germany, 1994.

L. J. Machlin, ed., *The Handbook of Vitamins*, Marcel Dekker, New York, 1991.

M. B. Sporn, A. B. Roberts, and D. S. Goodman, eds., *The Retinoids*, Vols. 1 and 2, Academic Press, Inc., Orlando, Fla., 1994.

Susan D. Van Arnum
Hoffmann-La Roche Inc.

VITAMIN B₁₂

Vitamin B₁₂ [68-19-9] (1,2) is the generic name for a closely related group of substances of microbial origin. Although the last of the vitamins to be characterized, its history is a long one, dating from 1824 when Combe (3) proposed the relationship of pernicious anemia, a disease characterized by defective (megoblastic) red blood cell formation, to disorders of the digestive system. Additional study of pernicious anemia, in particular the work of Addison (4), continued for over 100 years before Minot and Murphy (5) reported that a diet containing large quantities of raw liver restored the normal level of red blood cells in patients with pernicious anemia. This clinical breakthrough was based on the findings of Whipple and Robscheit-Robbins (6) that liver was of benefit in regeneration of blood in anemic dogs. For this work, Whipple, Minot, and Murphy were awarded the Nobel Prize in medicine and physiology in 1934.

In 1929, Castle (7) tied the work of Combe and Addison with that of Whipple, Minot, and Murphy by proposing that both an extrinsic factor and an intrinsic factor are involved in the control of pernicious anemia. The extrinsic factor, from food, is vitamin B₁₂. The intrinsic factor is a specific B₁₂-binding protein secreted by the stomach. This protein is required for vitamin B₁₂ absorption.

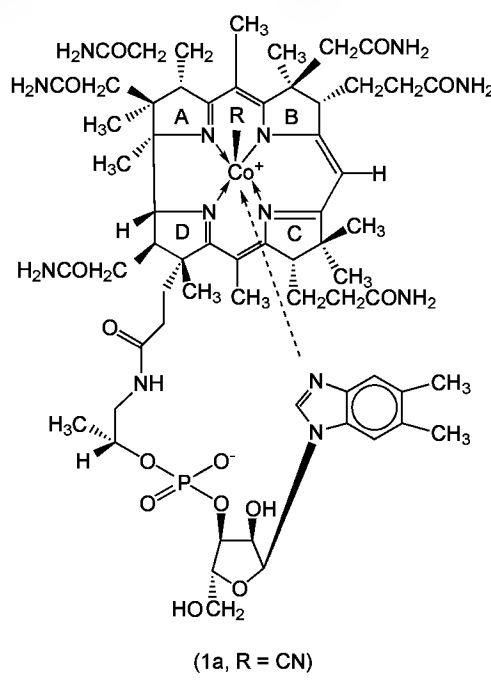
Work for the next 20 years focused on purification of the extrinsic factor from liver. The work was slow and tedious, because fractionation was guided by tests on pernicious anemia patients. The discovery (8) that *Lactobacillus lactis Dornier* requires liver extracts for growth greatly accelerated the isolation process. In 1948, groups at Merck (9) in the United States and Glaxo (10) in England reported isolation of vitamin B₁₂ (cyanocobalamin) as a crystalline, red pigment. The clinical efficacy of this material in the treatment of pernicious anemia was rapidly established (11).

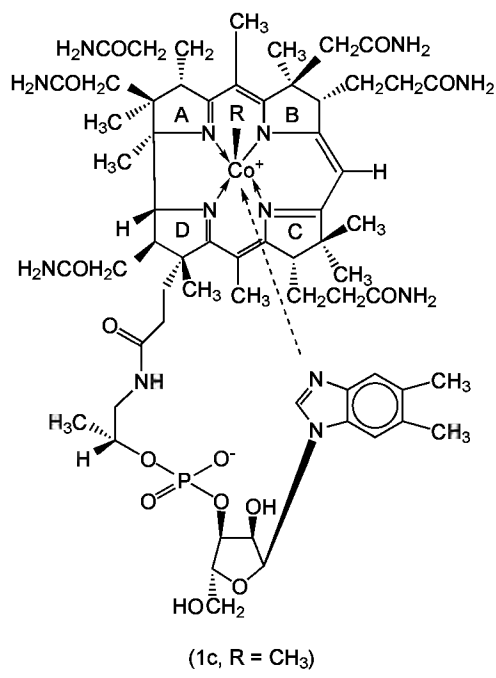
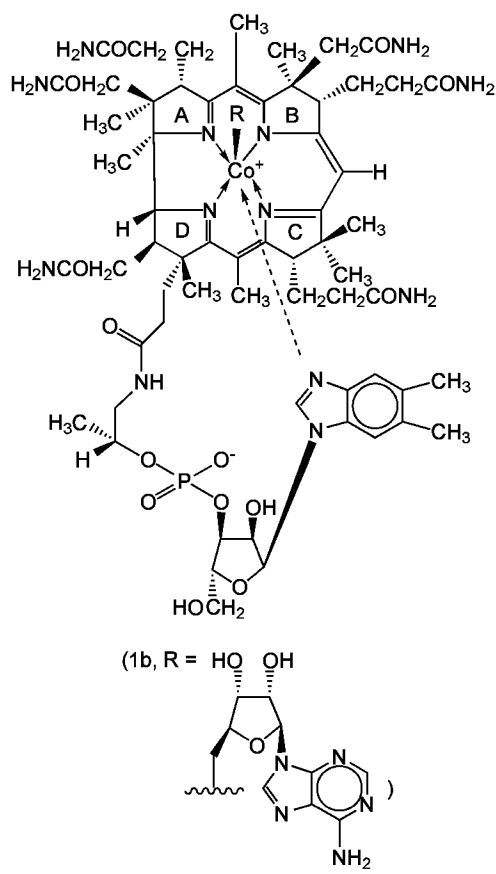
Parallel to the activities in the treatment of pernicious anemia were observations in the 1930s that most farm animals had a requirement for an unknown factor beyond the vitamins then known. The lack of this factor became apparent, eg, when chicks or pigs fed a diet with only vegetable protein evidenced slow growth rate and high mortality. It became apparent that the required factor, termed animal protein factor, was present in animal sources such as meat and tissue extracts, milk, whey, and cow manure. Subsequent to its isolation, it was rapidly shown that vitamin B₁₂ is the same as animal protein factor.

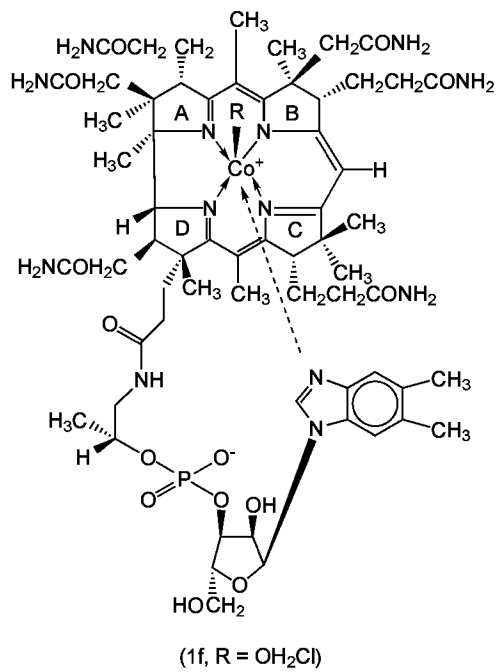
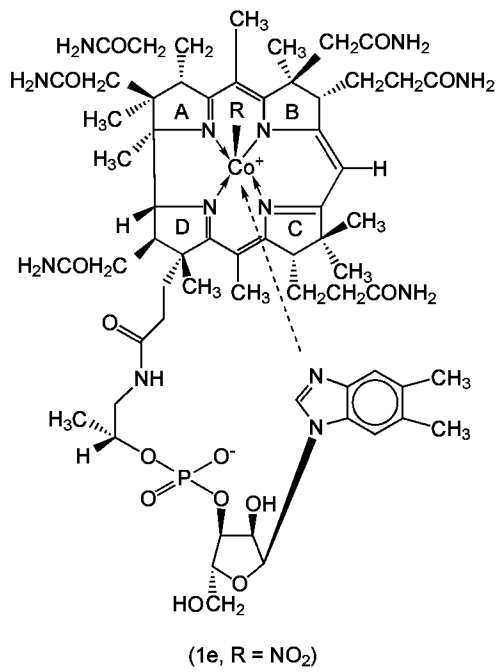
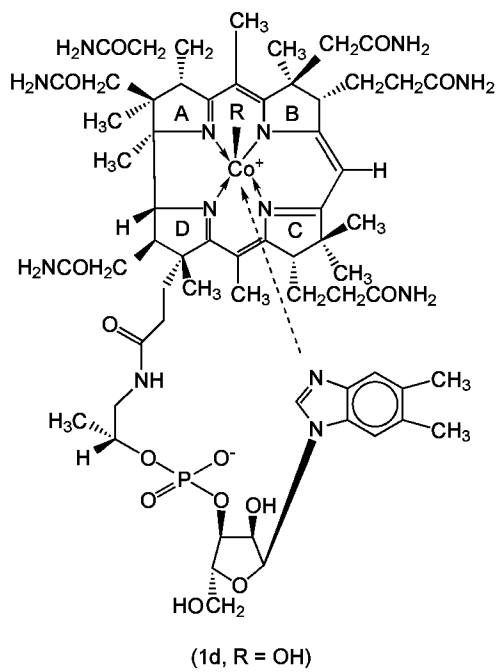
After its separation from liver extracts, vitamin B₁₂ was also isolated from cultures of *Streptomyces aureofaciens* (12). All vitamin B₁₂ sold commercially is produced by microbial fermentation.

Structure

The structure of the first isolated vitamin B₁₂, cyanocobalamin [68-19-9] (1a) is known to occur only sporadically, at best, in biological systems. Its isolation was an artifact resulting, probably, from use of charcoal containing cyanide in the purification process.





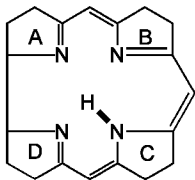


It is recognized that there are several important forms of vitamin B₁₂. The active coenzyme forms are adenosylcobalamin [13870-90-1] (coenzyme B₁₂ (1b)) and methylcobalamin [13422-55-4] (1c). These, along with hydroxocobalamin [13422-51-0] (vitamin B_{12a} (1d)), are the forms found in humans and other animals. Other forms of interest are nitrocobalamin (vitamin B_{12c} (1e)) and aquacobalamin chloride [13422-52-1] (vitamin B_{12b} (1f)). The primary commercial form is cyanocobalamin, due to its ease of isolation and purification as well as its stability.

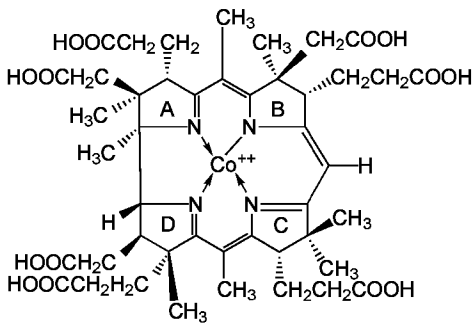
The IUPAC-IUB Commission on Biochemical Nomenclature (13) recommends that the term vitamin B₁₂ be used as the generic descriptor for all corrinoids exhibiting qualitatively the biological activity of cyanocobalamin. However, because of its commercial importance, cyanocobalamin is used interchangeably with vitamin B₁₂ herein.

Determination of the structure of vitamin B₁₂ was a slow, tedious process with the tools available in the 1940s. Despite extensive information gathered by degradation studies, the structures of cyanocobalamin and its coenzyme forms were only established beginning in 1955 by x-ray crystallography (14–16). The value of this work in the study of natural products was recognized by the award to Hodgkin of the Nobel Prize in chemistry in 1964.

Vitamin B₁₂ belongs to the class of molecules known as corrins. The core corrin structure (2) consists of four linked, partially saturated pyrrole rings. Corrin is a truncated form (no CH₂ between rings A and D) of the more common porphyrin skeleton. The corrin ring in vitamin B₁₂ is substituted in a highly regio- and stereospecific manner with eight methyl groups, three acetic acid chains, and four propionic acid chains, as in cobyrinic acid (3). The six conjugated double bonds of the corrin system give octahedral complexes with a number of metals. However, only complexes with cobalt exhibit vitamin B₁₂ activity.

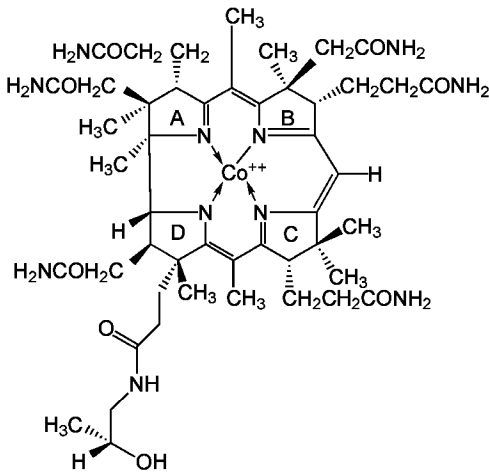


(2)



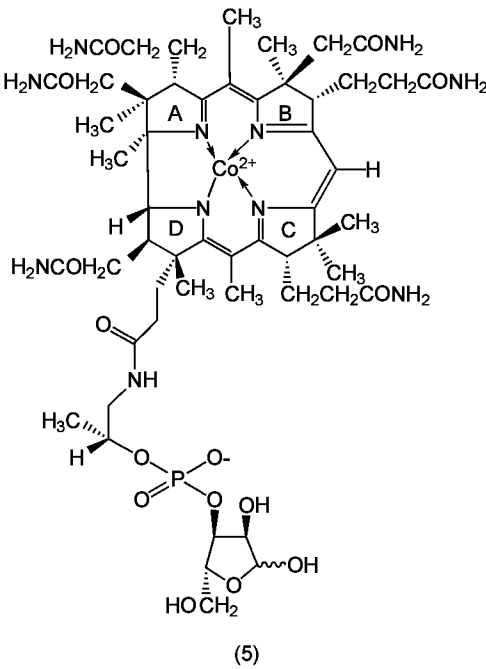
(3)

Cobrynic acid in which the propionic acid is amidated with 1-amino-2-propanol is known as cobinic acid, whereas the compound in which all other acids are primary amides is cobinamide (4).



(4)

Cobinic acid and cobinamide, which are linked to ribose 3-phosphate, are known as cobamic acid and cobamide (5), respectively.



The cobalt corrin complex is octahedral. The four nitrogens of the corrin ring occupy the four equatorial ligand positions in a virtually planar arrangement. The nucleotide formed from cobamide and 5,6-dimethylbenzimidazole provides a fifth ligand in the alpha, axial position of vitamin B₁₂. The group occupying the sixth, beta-position can vary substantially, as noted in structure (1), and dictates the compound name. Thus, when the ligand is cyanide, the compound name is cyanocobalamin; when methyl, methylcobalamin, etc.

Vitamin B₁₂ exists as a neutral complex. The beta-ligand and one nitrogen atom in the corrin ring each contribute a negative charge to Co(III). The nucleotide phosphate contributes the final negative charge. The structure of vitamin B₁₂ is highly organized and compact, as dictated by charge neutralization and the ligand sphere of the cobalt atom.

Many analogues of vitamin B₁₂ are known. These occur either naturally or are obtained by synthesis, degradation of the vitamin, or directed biosynthesis. Most of these compounds have little or no biological activity in animals. However, many have significant microbiological activity. Among the analogues found are those in which the 5,6-dimethylbenzimidazole is replaced by another (or no) base. These include 2-methyl adenine (Factor A), no base (Factor B), guanine (Factor C), 5-hydroxybenzimidazole (Factor III), 5-methoxybenzimidazole (Factor III_m) and adenine (pseudovitamin B₁₂) (17–20).

Occurrence

For many years, it was thought that the occurrence of vitamin B₁₂ was limited to animal tissues and bacteria, with all of the material originating from bacterial sources. More recently, the presence of vitamin B₁₂ and or vitamin B₁₂-like activity at low levels in plants has been recognized (21–24). Nonetheless, the largest portion of the vitamin B₁₂ requirement is met by consumption of animal tissues and from microorganisms in the animal's digestive tract. Herbivorous animals satisfy their vitamin B₁₂ needs by absorption of material produced by rumenal or intestinal flora. In humans and carnivorous omnivorous animals, intestinal production of vitamin B₁₂ occurs but at an insufficient level. As a result, vitamin B₁₂ levels are maintained by consumption of foods rich in vitamin B₁₂ such as liver, heart, and kidney. Egg yolk and some fish and shellfish are also good sources. Vitamin B₁₂ supplements, both oral and parenteral, are also available. Dietary sources of foodstuffs are provided in Table 1 (25–27).

Table 1. Dietary Sources of Vitamin B₁₂

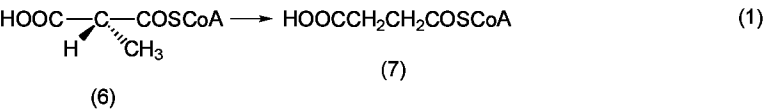
>50 µg/100 g	5–50 µg 100 g	<5 µg/100 g
liver	liver	beef (lean)
lamb	rabbit	lamb
beef	chicken	pork
calf	kidney	chicken
pork	rabbit	egg (whole)
kidney	beef	cheese
lamb	fish	american
brain	sardines	swiss
beef	salmon	milk (cow)
	herring	fish
	heart	cod
	beef	flounder
	rabbit	haddock
	chicken	sole
	egg yolk	halibut
	clams	swordfish
	oysters	tuna
	crabs	mackerel
		lobster
		scallop
		shrimp
		green vegetables

Biochemical Functions

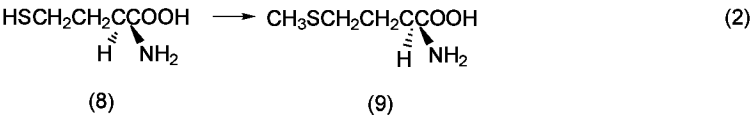
Methylcobalamin and adenosylcobalamin are the two coenzyme forms of vitamin B₁₂ in animals and humans. Each is involved in the catalysis of a specific

transformation. In humans, it appears that there are only two enzymes requiring vitamin B₁₂ as an essential coenzyme although many other, particularly bacterial enzyme systems also require a vitamin B₁₂ coenzyme (2,28).

Adenosylcobalamin (coenzyme B₁₂) is required in a number of rearrangement reactions; that occurring in humans is the methylmalonyl-CoA mutase-mediated conversion of (R)-methylmalonyl-CoA (6) to succinyl-CoA (7) (eq. 1). The mechanism of this reaction is poorly understood, although probably free radical in nature (29). The reaction is involved in the catabolism of valine and isoleucine. In bacterial systems, adenosylcobalamin drives many 1,2-migrations of the type exemplified by equation 1 (30).



Methylcobalamin is involved in a critically important physiological transformation, namely the methylation of homocysteine (8) to methionine (9) (eq. 2) catalyzed by N⁵-methyltetrahydrofolate homocysteine methyltransferase. The reaction sequence involves transfer of a methyl group first from N⁵-methyltetrahydrofolate to cobalamin (yielding methylcobalamin) and thence to homocysteine. Once again, the intimate details of the reaction are not well known (31). Demethylation of tetrahydrofolate to tetrahydrofolic acid is a step in the formation of thymidine phosphate, in turn required for DNA synthesis. In the absence of the enzyme, excess RNA builds up in red blood cells.



Homocysteine has been identified as an independent risk factor for atherosclerosis (32) and thus metabolic control over homocysteine levels has major health implications.

Deficiency. Macrocytic anemia, megaloblastic anemia, and neurological symptoms characterize vitamin B₁₂ deficiency. Alterations in hematopoiesis occur because of the high requirement for vitamin B₁₂ for normal DNA replication necessary to sustain the rapid turnover of the erythrocytes. Abnormal DNA replication secondary to vitamin B₁₂ deficiency produces a defect in the nuclear maturational process of committed hematopoietic stem cells. As a result, the erythrocytes are either morphologically abnormal or die during development.

Neurological symptoms result from demyelination of the spinal cord and are potentially irreversible. The symptoms and signs characteristic of a vitamin B₁₂ deficiency include paresthesia of the hands and feet, decreased deep-tendon reflexes, unsteadiness, and potential psychiatric problems such as moodiness, hallucinations, delusions, and psychosis. Neuropsychiatric disorders sometimes develop independently of the anemia, particularly in elderly patients. Visual loss may develop as a result of optic atrophy.

Clinical manifestation of vitamin B₁₂ deficiency is usually a result of absence of the gastric absorptive (intrinsic) factor. Dietary deficiency of vitamin B₁₂ is uncommon and may take 20 to 30 years to develop, even in healthy adults who follow a strict vegetarian regimen. An effective enterohepatic recycling of the vitamin plus small amounts from bacterial sources and other contaminants greatly minimizes the risk of a complete dietary deficiency. Individuals who have a defect in vitamin B₁₂ absorption, however, may develop a deficiency within three to seven years.

Dietary deficiency in the absence of absorption defects can be effectively reversed with oral supplementation of 1 μm of vitamin B₁₂ daily. If deficiency is related to a defect in vitamin absorption, daily doses of 1 μg administered subcutaneously or intramuscularly are effective (33). However, a single intramuscular dose of 100 μg of cobalamin once per month is adequate in patients with chronic gastric or ileal damage. Larger doses are generally rapidly cleared from the plasma into the urine and are not effective unless the patient demonstrates poor vitamin retention.

Requirement. A daily intake of 1 μg should cover the daily loss of vitamin and maintain an adequate body pool. The RDA (34), however, has been established at 2 μg day to cover metabolic variation among individuals and to ensure normal serum concentrations and adequate pool sizes (Table 2).

Table 2. The 1989 RDA for Vitamin B₁₂

Group	Age, yr	Vitamin B ₁₂ , μg
RDA		
infants	0–0.5	0.3
	0.5–1.0	0.5
children	1–3	0.7
	4–6	1.0
	7–10	1.4
males	11–51 +	2.0
females	11–51 +	2.0
	pregnant	2.2
	lactating	2.6
U.S. RDI (daily value)		
infants and children <4 yr		3.0
adults and children >4 yr		6.0
pregnant or lactating women		8.0

Smaller pool sizes with normal serum B₁₂ levels may be maintained with dietary intakes below 1 μg. However, more substantial pool sizes are considered advantageous as protection against the development of pernicious anemia, which may occur in advanced age; achlorhydria becomes more common after age 60, resulting in compromised absorption of vitamin B₁₂.

In general, maternal stores of vitamin B₁₂ are considered adequate to meet the demands of pregnancy.

Safety. No toxicity has been associated with acute or chronic intakes of vitamin B₁₂ in doses of 100 and 1 mg, respectively. Vitamin B₁₂ absorption is both limited and affected by vitamin status. Therefore, absorption is reduced with improved status, lessening the risk of toxicity.

Cobalamin should be administered parenterally by the intramuscular or subcutaneous route. Isolated cases of anaphylaxis have been reported with intravenous administration.

Metabolism

Absorption. An absorption mechanism capable of handling between 1.5 to 3.0 μg of vitamin B₁₂ is responsible for most of the intestinal

absorption. This mechanism includes the binding of vitamin B₁₂ to a specific transport protein (intrinsic factor) and the presence of specific membrane-bound receptors on the ileal cell surface in the small intestine. In addition, there is a second mechanism for the absorption of vitamin B₁₂ by diffusion. This mechanism can provide a physiologically significant source of the vitamin when delivered in pharmacological dosages.

Food vitamin B₁₂ appears to bind to a salivary transport protein referred to as the R-protein, R-binder, or haptocorrin. In the stomach, R-protein and the intrinsic factor competitively bind the vitamin. Release from the R-protein occurs in the small intestine by the action of pancreatic proteases, leading to specific binding to the intrinsic factor. The resultant complex is transported to the ileum where it is bound to a cell surface receptor and enters the intestinal cell. The vitamin B₁₂ is then freed from the intrinsic factor and bound to transcobalamin II in the enterocyte. The resulting complex enters the portal circulation.

Transport. Transcobalamin II delivers the absorbed vitamin B₁₂ to cells and is the primary plasma vitamin B₁₂-binding transport protein. It is found in plasma, spinal fluid, semen, and extracellular fluid. Many cells, including the bone marrow, reticulocytes, and the placenta, contain surface receptor sites for the transcobalamin II–cobalamin complex.

Other plasma vitamin B₁₂ proteins, transcobalamines I and III, appear to have primarily a storage function and only a lesser role in transport.

Tissue Uptake and Storage. Cell surface receptors take up the transcobalamin II–cobalamin complex, which is internalized into endosomes. The complex is dissociated and the transcobalamin II released. The mechanism by which cobalamin leaves the endosome is uncertain.

The liver is the principal site of vitamin B₁₂ storage, containing between 50 and 99% of the total body pool. The average total body pool is estimated to range between 2.0 and 5.0 mg. Higher storage values reported probably reflect noncobalamin analogues as well as cobalamin. The main storage form of vitamin B₁₂ appears to be coenzyme B₁₂. The primary circulating form in the plasma appears to be methylcobalamin.

Metabolism and Mobilization. On entry of vitamin B₁₂ into the cell, considerable metabolism of the vitamin takes place. Co(III)cobalamin is reduced to Co(I)cobalamin, which is either methylated to form methylcobalamin or converted to adenosylcobalamin (coenzyme B₁₂). The methylation requires methyl tetrahydrofolate.

Approximately 0.05 to 0.2% of vitamin B₁₂ stores are turned over daily, amounting to 0.5–8.0 µg, depending on the body pool size. The half-life of the body pool is estimated to be between 480 and 1360 days with a daily loss of vitamin B₁₂ of about 1 µg. Consequently, the daily minimum requirement for vitamin B₁₂ is 1 µg. Three micrograms (3.0 µg) vitamin B₁₂ are excreted in the bile each day, but an efficient enterohepatic circulation salvages the vitamin from the bile and other intestinal secretions. This effective recycling of the vitamin contributes to the long half-life. Absence of the intrinsic factor interrupts the enterohepatic circulation. Vitamin B₁₂ is not catabolized by the body and is, therefore, excreted unchanged. About one-half of the vitamin is excreted in the urine and the other half in the bile.

Properties

Table 3 lists some of the physical and chemical properties of vitamin B₁₂ (1,35). Crystalline vitamin B₁₂ is stable in air and is not affected by moisture. The anhydrous compound, however, is very hygroscopic, and when exposed to moist air may absorb about 12% of water.

Table 3. Physical and Chemical Properties of Vitamin B₁₂ (Cyanocobalamin)^a

Property	Characteristic
appearance	crystalline
color	dark red
mol wt	1355.42
empirical formula	C ₃ H ₈₈ CoN ₁₄ O ₁₄ P
mp, °C	darkens 210–220; does not melt <300
specific rotation	α_{D}^{25} 59 ± 9° (diluted aqueous)
uv absorption	278,361,550 nm (water)
pH (aq) ^b	neutral
solubility %	
H ₂ O	1.25
alcohol	2.03
acetone	insoluble
ether	insoluble

^a Properties given for the anhydrous compound.
^b Vitamin B₁₂ exhibits maximum stability between pH 4.5 and 5.0.

Aqueous solutions of vitamin B₁₂ at pH 4.0 to 7.0 show no decomposition during extended storage at 25°C. For optimum stability at elevated temperatures, solutions should be adjusted to pH 4.0 to 4.5. Aqueous solutions in this pH range may be autoclaved for 20 min at 120°C without significant decomposition.

Redox Reactions. Critical to the function of cobalamins as enzyme cofactors is the ability of the cobalt atom to exist in the Co(III) and Co(I) oxidation states. Chemically, Co(II) forms are also known. Each oxidation state has different ligand-accepting abilities. The chemical (1a, 1d–1f) and coenzyme (1b,1c) forms of cobalamin are trivalent. These compounds are readily reduced to Co(II) and Co(I)cobalamins. The redox potentials for the aquacobalamin to Co(II)cobalamin and Co(II)cobalamin to Co(I)cobalamin couples are 0.04 and 0.85 V, respectively (36). Chemical reduction of aquacobalamin to Co(II)cobalamin is effected by thiols or carbon monoxide. Cobalt(II)cobalamin contains a single unpaired electron in the 3d_{xy} orbital of the cobalt ion; one of the two axial positions is unoccupied and thus this material is five-coordinate. Chemically, Co(I)cobalamin is obtained with strong reducing agents, eg, NaBH₄ or zinc and NH₄Cl. It is a powerful reducing agent, and reacts rapidly with oxygen and reduces protons to hydrogen. It is, therefore, unstable in aqueous acid solution, less so in neutral or basic aqueous solution. It is frequently used for the preparation of organocobalamins, eg, adenosylcobalamin and methylcobalamin. Co(I)cobalamin contains two electrons in the 3d_{xy} orbital. It has been postulated that the coordination number for the cobalt is four and that both axial ligands are vacant (37).

Exchange of Axial Ligands. Many ligand-exchange reactions involve groups in which the coordination to the metal is through nitrogen (NH₃,N₃), oxygen (H₂O,OH), sulfur (SH,SO₂), halogen, or carbon (CN,CH₃). Important reactions involve displacement of the heterocyclic base from the alpha-coordination position by a solvent, usually water. This displacement occurs in acidic solution and results from the protonation of the heterocyclic base. The protonation is associated with a characteristic change in the spectrum. The pK_a for the base-on /base-off equilibrium depends on the nature of the beta-ligand. Displacement of H₂O, adenosyl, or methyl from cobalt by cyanide has also been studied. In the presence of cyanide ion, aquacobalamin and adenosylcobalamin are converted to cyanocobalamin. In contrast, methylcobalamin and other alkyl corrinoids are stable in the presence of 0.1 M cyanide in the dark (37,38). The equilibrium, aquacobalamin + H⁺ (pK_a 6.9–7.8), is not a pure ligand exchange but only a ligand modification. The cobalt-bound water is acidic and reversibly loses a proton at neutral pH.

Chemical Reactions. A wealth of information exists on the chemistry of vitamin B₁₂ (1,39,40). Much of this chemistry was established during

the studies leading to structure determination, partial synthesis, and the total synthesis of vitamin B₁₂. Vitamin B₁₂ is slowly decomposed by ultraviolet or strong visible light. By controlled irradiation with visible light, cyanide may be selectively liberated from cyanocobalamin without destruction of the cobalamin structure, but long exposure to light results in complete inactivation of the vitamin. Vitamin B₁₂ is also inactivated by treatment with strong acids or bases.

One development involves the use of vitamin B₁₂ to catalyze chemical, in addition to biochemical processes. Vitamin B₁₂ derivatives and B₁₂ model compounds (41,42) catalyze the electrochemical reduction of alkyl halides and formation of C–C bonds (43,44), as well as the zinc–acetic acid-promoted reduction of nitriles (45), alpha, beta-unsaturated nitriles (46), alpha, beta-unsaturated carbonyl derivatives and esters (47,48), and olefins (49). It is assumed that these reactions proceed through intermediates containing a Co–C bond which is then reductively cleaved.

Analysis

Specifications. Cyanocobalamin is the commercial form of vitamin B₁₂. It is sold under the following tradenames (35):

Anacobin	Ducobee
Antipemycin	Duodecibin
Bedoce	Embiol
Bedodeka	Emociclina
Bedoz	Eritrone
Behepan	Erycytol
Berubi	Erythrotin
Berubigen	Euhaemon
Betalin-12	Fresmin
Betolvex	Hemo-B-Doze
Bevatine-12	Hemomin
Bevidox	Hepagon
Bexii	Hepavis
Bexil	Hepcovite
Biocobalamine	Hydroxamin
Biocres	Hydroxobase
Bitevan	Macrabin
B-Telve	Megabion (Indian)
B-Twelv	Megalovel
Byladoce	Milbedoce
Claretin-12	Millevit
Cobalin	Nagravon
Cobamin	Normocytin
Cobamine	Peraemon
Cobione	Pernaevit
Covit	Pernipur
Crystamin	Plecyamin
Cycobemin	Poyamin
Cycolamin	Redamina
Cykobeminet	Redisol
Cytacon	Rhodacryst
Cytamen	Rubesol
Cytobion	Rubivitan
Distivit (B ₁₂ peptide)	Rubramin
Dobetin	Rubripca
Docemine	Rubrocitol
Docibin	Sytobex
Docigram	Vitalt
Docivit	Vibisone
Dodecabee	Virubra
Dodecavite	Vitarubin
Dodex	Vita-Rubra
	Vitral

Specifications are found in the *Codex* for food use (50) and in the USP (51) for pharmaceutical use.

Analytical Methodology. Vitamin B₁₂ can be determined by microbiological, radioisotope dilution, spectrophotometric, chemical, or biological methods employing animals (52–54). Microbiological assays involve the extraction and stabilization of vitamin B₁₂ from the food, feed, or pharmaceutical matrix prior to assay. The official method of the AOAC (55) accomplishes this by autoclaving samples in a phosphate–citric acid buffer containing metabisulfite, whereas the British Analytical Methods Committee (56) recommends extraction with aqueous cyanide solution at pH 4.6–5.0 in a boiling water bath. The AOAC extraction procedure was found to yield somewhat higher results than the British Analytical Methods Committee procedure (57).

The AOAC assay is based on the graded growth of the bacterium *Lactobacillus leichmannii* ATCC 7830 in a medium containing all required growth factors but vitamin B₁₂. Results are obtained by determining the transmittance of the sample and standard tubes after 16–24 h incubation at 30–40°C or by titrating the acid produced after 72 h incubation. The British Analytical Methods Committee assay employs the protozoan *Ochromonas malhamensis* ATCC 11532 in an assay based on the graded growth of the protozoan. The assay incubation period varies from 3–6 days and degree of growth is measured turbidimetrically. Although the *L. leichmannii* assay is claimed to be less specific than the *O. malhamensis* assay it has several advantages over the latter. The AOAC procedure has a shorter incubation period and the assay is set up in test tubes, whereas the *O. malhamensis* assay is set up in 25-mL micro-Fernbach flasks, the assay medium is less complex, and samples whose turbidity or color would interfere in a turbidimetric assay can be analyzed by the titrimetric assay. Derivatives of deoxynucleic acid that stimulate the growth of *L. leichmannii* in the absence of vitamin B₁₂ can be measured after destroying the vitamin B₁₂ by autoclaving the sample at pH 11–12 and determining the residual activity. The bacteria *Lactobacillus lactis* Dorner ATCC 8000 and *Escherichia coli* ATCC 10799 and 14169 as well as the protozoan *Euglena gracilis* ATCC 12716 have been used also for the determination of vitamin B₁₂.

Radioisotope dilution assays are based on the principle of competition between radioactive labeled (⁶⁷Co) vitamin B₁₂ and cobalamins extracted from matrices for binding sites on the intrinsic factor (a glycoprotein). Binding is in proportion to the concentration of the radioactive and nonradioactive B₁₂ with the concentration of intrinsic factor as the limiting factor. Free cobalamins are separated from those bound on the intrinsic factor by absorption

onto treated charcoal and the amount of free-labeled vitamin B₁₂ is determined. Vitamin B₁₂ content of the sample is determined from a standard curve. Results obtained by a radioisotopic dilution method are very similar to those obtained by the AOAC microbiological assay (58).

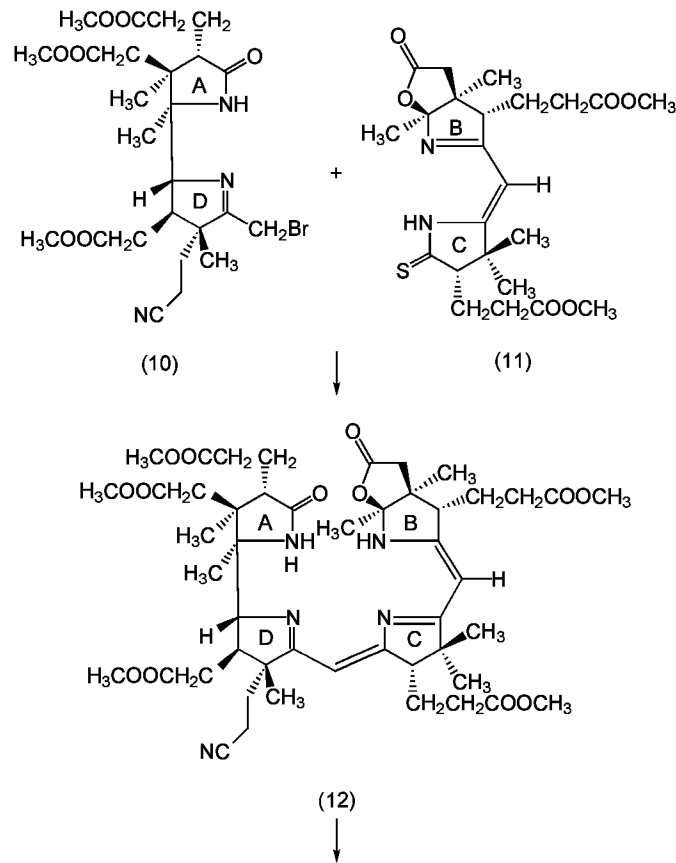
Spectrophotometric determination at 550 nm is relatively insensitive and is useful for the determination of vitamin B₁₂ in high potency products such as premixes. Thin-layer chromatography and open-column chromatography have been applied to both the direct assay of cobalamins and to the fractionation and removal of interfering substances from sample extracts prior to microbiological or radioassay. Atomic absorption spectrophotometry of cobalt has been proposed for the determination of vitamin B₁₂ in dry feeds. Chemical methods based on the estimation of cyanide or the presence of 5,6-dimethylbenzimidazole in the vitamin B₁₂ molecule have not been widely used.

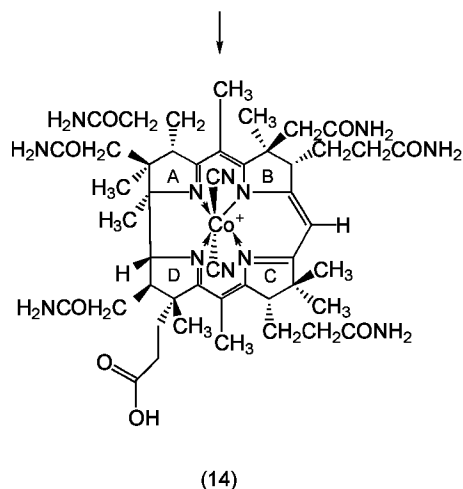
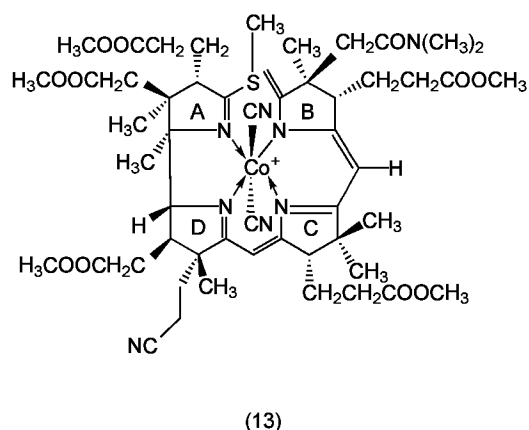
Various aspects of the chromatography of vitamin B₁₂ and related corrinoids have been reviewed (59). A high performance liquid chromatographic (hplc) method is reported to require a sample containing 20–100 µg cyanocobalamin and is suitable for premixes, raw material, and pharmaceutical products (60).

Bioassays are based on the growth response of vitamin-depleted rats or chicks to graded amounts of vitamin B₁₂ added in the diet. These assays are not specific for vitamin B₁₂ because factors, other than vitamin B₁₂ present in biological materials, produce a growth response. Because coenzyme B₁₂, a primary form of natural vitamin B₁₂, is light sensitive, assays should be carried out in subdued light.

Synthesis

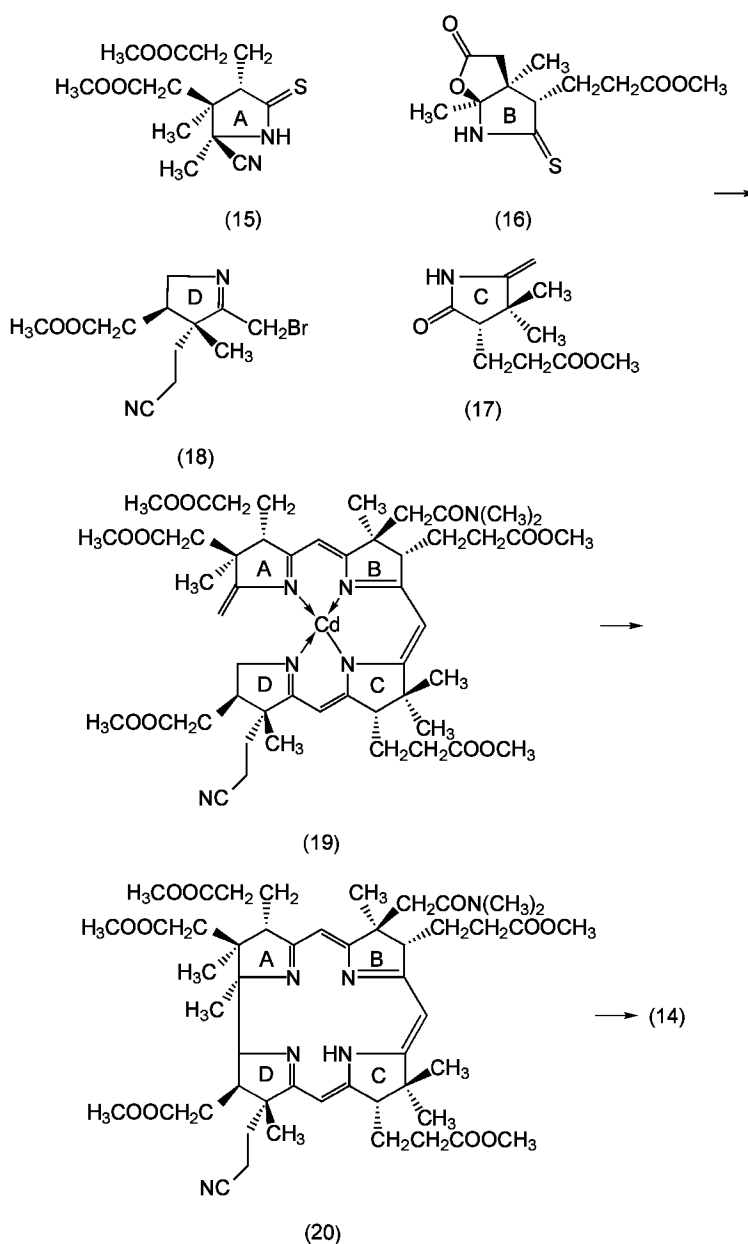
The achievements in total synthesis of organic compounds in recent years are perhaps nowhere better illustrated than in the synthesis of vitamin B₁₂ by the groups of Woodward (61–63) and Eschenmoser (64–69) in a collaborative effort (70,71). The work has been reviewed (1,72,73). The real value of the synthesis lies in the synthetic methodology developed in the course of the work. During the synthesis of the A–D fragment, the Woodward-Hoffmann rules for the conservation of orbital symmetry (74) were developed. For this work, Hoffmann shared the Nobel Prize in chemistry in 1981 with Fukui. Woodward was awarded the Nobel Prize in chemistry in 1965 for the synthesis and structure elucidation of natural products, including the work leading up to the synthesis of vitamin B₁₂. Thus, the history of vitamin B₁₂ is intimately connected with the awarding of four Nobel Prizes to date.





The core of the first synthesis of vitamin B₁₂ involved condensation of the A–D ring fragment (10) with the B–C fragment (11). The former compound was obtained by the Harvard group in ca 35 steps, including a classical resolution, from 3-methoxyaniline and camphor.

The synthesis of the second fragment (11) by the group at the ETH in Zurich occurred in 18 steps from camphorquinone (ring C) and *trans*-3-methyl-4-oxopentenoic acid (ring D). Base-catalyzed alkylation of (11) with (10) yielded a thioiminoether that underwent sulfide contraction upon treatment with acid to give the tetrapyrrole (12). Manipulation of substituents and introduction of the cobalt atom gave complex (13). The stereo-organizing template effect of the complex allowed base-catalyzed cyclization to complete the corrin ring system. Functional group exchange and addition of the final methyl group then yielded cobyrinic acid (14). This material had previously been converted to vitamin B₁₂ (75) and thus its obtention by total synthesis completed the synthesis.



A second synthesis of cobyrinic acid (14) involves photochemical ring closure of an A-D secocorrinoid. Thus, the Diels-Alder reaction between butadiene and *trans*-3-methyl-4-oxopentenoic acid was used as starting point for all four ring A-D synthons (15–18). These were combined in the order B + C → BC + D → BCD + A → ABCD. The resultant cadmium complex (19) was photocyclized in buffered acetic acid to give the metal-free corrinoid (20). A number of steps were involved in converting this material to cobyrinic acid (14).

The total syntheses have yielded cobyrinic acid and thence cyanocobalamin. Routes to other cobalamins, eg, methylcobalamin and adenosylcobalamin, are known (76–79). One approach to such compounds involves the oxidative addition of the appropriate alkyl halide (eg, CH₃I to give methylcobalamin) or tosylate (eg, 5'-*p*-tosyladenosine to yield adenosylcobalamine) to cobalt(I)alamin.

The complexity of the vitamin B₁₂ molecule makes it extremely unlikely that total synthesis will ever be employed for preparation of commercial quantities.

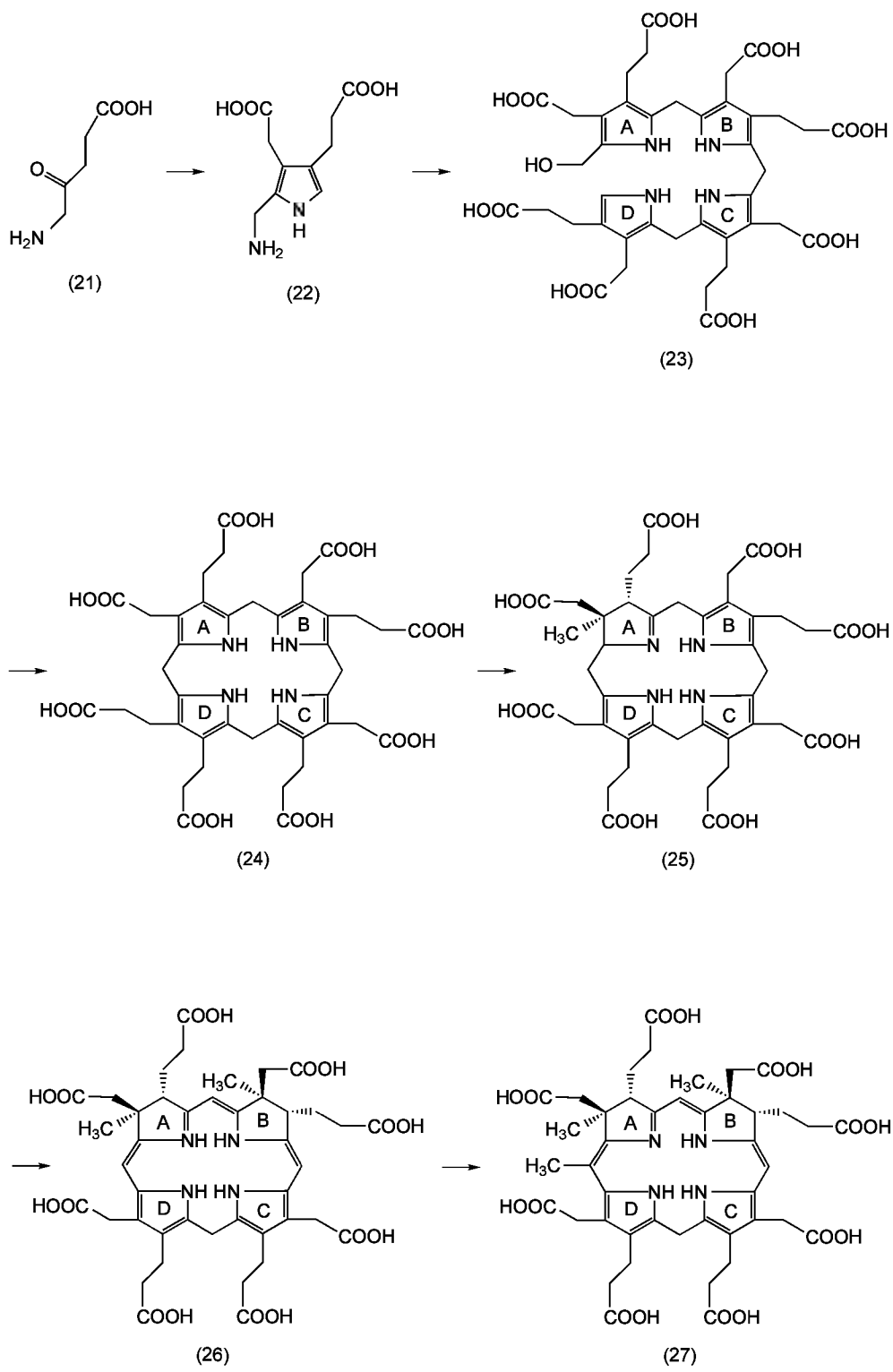
Biosynthesis

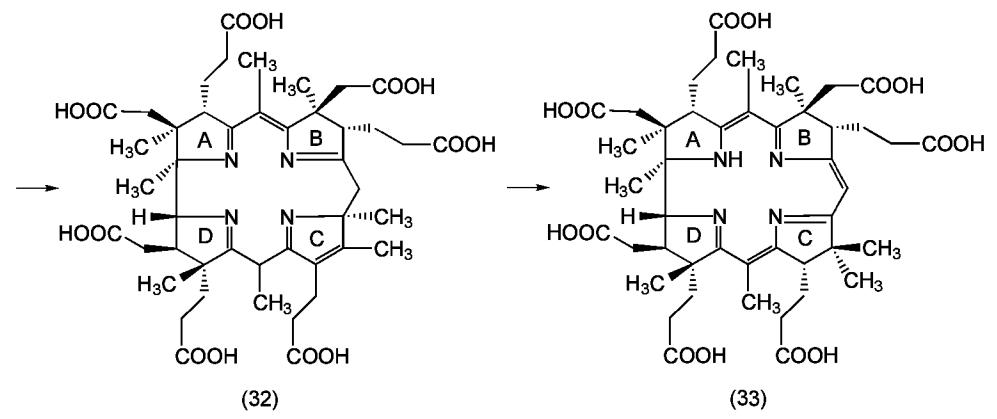
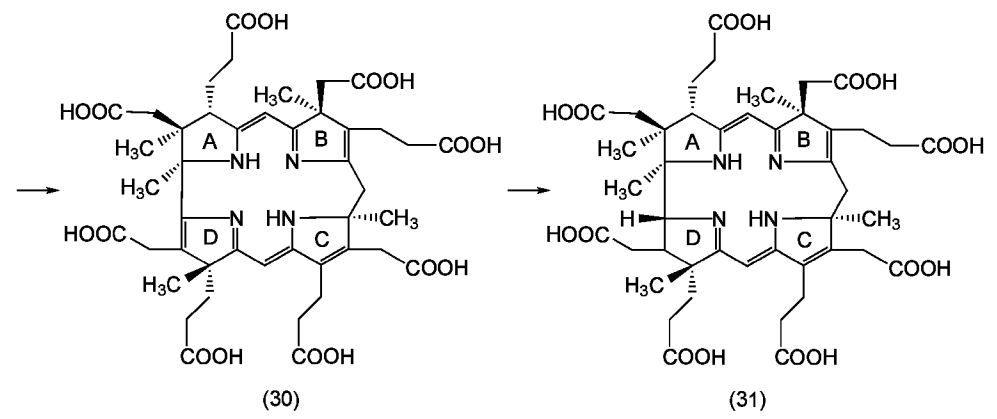
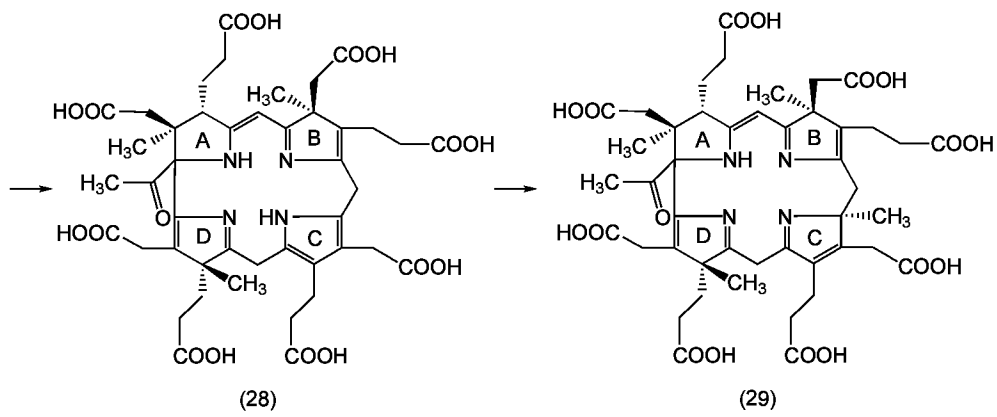
The study of the biosynthesis of vitamin B₁₂ is a saga whose resolution, due primarily to Battersby (80–83) and Scott (84,85), required an effort on the same magnitude as the total synthesis. It was only when recent molecular biology tools became available to complement enzymology, isotopic labeling, chemical synthesis, and spectroscopy that solution of this problem became possible.

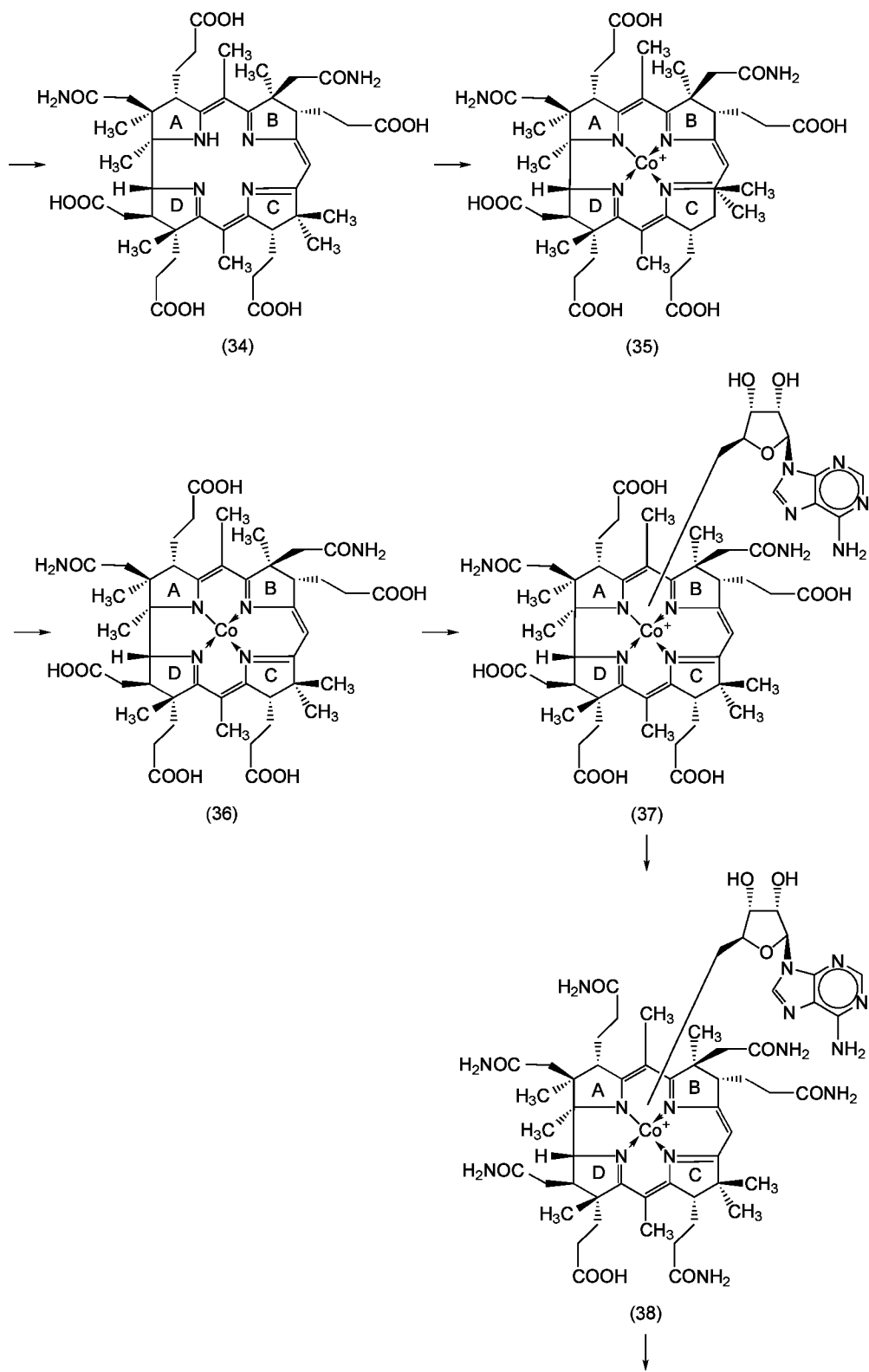
The biosynthesis of vitamin B₁₂ in the species *Pseudomonas denitrificans* has received the most attention due to its importance as a commercial source of the vitamin. The 22 genes involved have been identified and the sequence of the resulting proteins described (83). The sequence is initiated from 5-aminolevulinic acid (21) which is biosynthesized from succinyl-CoA (7) and glycine. Condensation of two molecules of 5-aminolevulinic acid yields porphobilinogen (22), which is tetramerized as a linear, head-to-tail arrangement to give hydroxymethylbilane (23). Ring closure and rearrangement (formally on exchange of acetate and propionate substituents in ring C) gives uroporphyrinogen III (uro'gen III) (24). Uro'gen III is a key biosynthetic branch point, leading not only to vitamin B₁₂ but also to chlorophyll and heme.

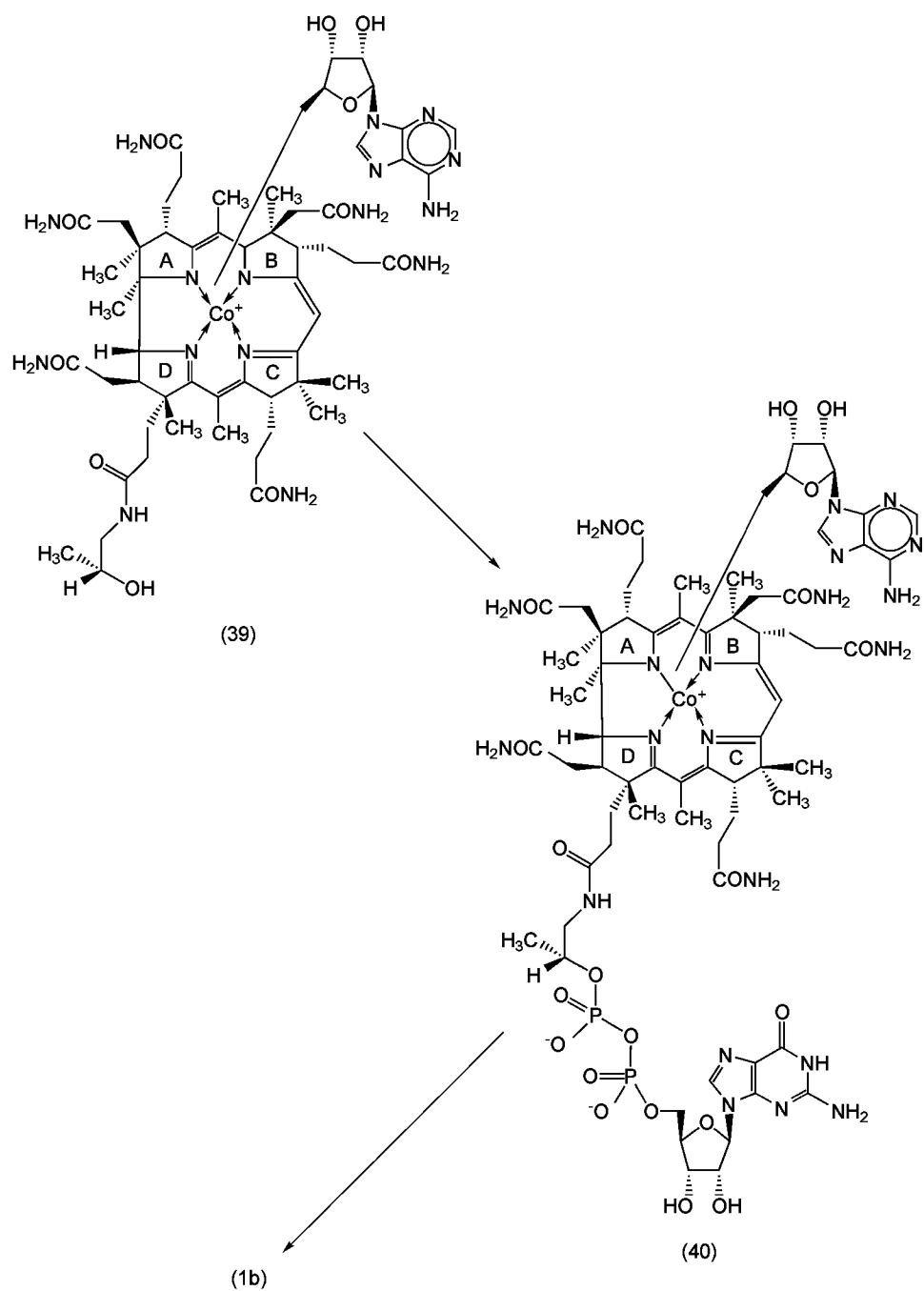
The sequence to vitamin B₁₂ proceeds by the introduction of the eight methyl groups, punctuated by ring contraction to the corrin, an oxidation and reduction, and a methyl migration. Thus, the first three methylations lead stereospecifically to precorrin-1 (25) and thence by precorrin-2 (26) to precorrin-3A (precorrin-3) (27). Oxidation, ring contraction to the corrin, and introduction of the fourth methyl group give precorrin-4 (28). Further methylation leads to precorrin-5 (29). Loss of the acetyl group created by ring contraction as acetate occurs with methylation to give precorrin-6A (precorrin-6x) (30), which is reduced to precorrin 6B (precorrin-6y) (31). Decarboxylation of the ring C acetic acid subsequent to addition of the last two methyl groups (precorrin-8x) (32) and methyl migration gives hydrogenobyrrinic acid (33).

Introduction of the cobalt atom into the corrin ring is preceded by conversion of hydrogenobyrrinic acid to the diamide (34). The resultant cobalt(II) complex (35) is reduced to the cobalt(I) complex (36) prior to adenosylation to adenosylcobyrinic acid *α,ε*-diamide (37). Four of the six remaining carboxylic acids are converted to primary amides (adenosylcobyrinic acid) (38) and the other amidated with (R)-1-amino-2-propanol to provide adenosylcobinamide (39). Completion of the nucleotide loop involves conversion to the monophosphate followed by reaction with guanosyl triphosphate to give diphosphate (40). Reaction with α -ribazole 5'-phosphate, derived biosynthetically in several steps from riboflavin, and dephosphorylation completes the synthesis.









The biosynthetic sequence for other aerobic bacteria appears, where known, to be similar to that in *Pseudomonas denitrificans* although the genes involved, and thus the enzymes, exhibit differences.

In anaerobic (or more correctly, almost anaerobic or microaerophilic) bacteria such as *Propionibacterium shermanii*, a fundamental difference occurs in that the cobalt atom is introduced at a much earlier stage, possibly to precorrin-2 or precorrin-3A.

Manufacture

As noted above, all vitamin B₁₂ is produced by microbial fermentation. A partial list of microorganisms that synthesize vitamin B₁₂ under appropriate conditions follows. Most strains, in their wild state, produce less than 10 mg L vitamin B₁₂, although a few approach 40 mg L. The organisms are both aerobes and anaerobes. The carbon requirements in the fermentations are satisfied from sources as wide ranging as hydrocarbons, methanol, and glucose.

- | | |
|--|--|
| <i>Arthrobacter hyalinus</i>
<i>Bacillus megaterium</i>
<i>Butyrivacterium rettgeri</i>
<i>Clostridium sticklandii</i>
<i>Clostridium tetanomorphum</i>
<i>Clostridium thermoaceticum</i> | <i>Nocardia rugosa</i>
<i>Propionibacterium arabinosum</i>
<i>Propionibacterium freudenreichii</i>
<i>Propionibacterium pentosaceum</i>
<i>Propionibacterium peterssoni</i>
<i>Propionibacterium shermanii</i>
<i>Propionibacterium technicum</i>
<i>Propionibacterium vannielli</i>
<i>Protaminobacter ruber</i>
<i>Pseudomonas denitrificans</i>
<i>Rhizobium meliloti</i>
<i>Rhodopseudomonas capsulata</i>
<i>Rhodopseudomonas sphaeroides</i>
<i>Strigomonas oncopelti</i>
<i>Streptomyces aureofaciens</i>
<i>Streptomyces griseus</i>
<i>Streptomyces olivaceus</i> |
| <i>Corynebacterium and Rhodopseudomonas</i>
<i>Critibidia fasciculata</i>
<i>Methanobacterium arbophilicum</i>
<i>Methanobacterium formicicum</i>
<i>Methanobacterium ruminantium</i> | |
| <i>Methanobacterium thermoautotrophicum</i>
<i>Methanosarcina barkeri</i>
<i>Micromonospora purpurea</i>
<i>Nocardia gardneri</i> | |

Commercial Production. Vitamin B₁₂, as cyanocobalamin, is produced by several companies. The market is dominated, however, by two

French firms, Rh  ne-Poulenc and, to a lesser extent, Roussel-Uclaf. Smaller amounts are produced in Japan by Nippon Petrochemical, in Hungary by Medimpex-Richter, and by minor producers in several other countries. Earlier manufacturers, particularly Merck (U.S.) and Glaxo (U.K.), have exited the market. Although estimates vary, it appears that ca 10,000 kg yr of vitamin B₁₂ is produced (1).

The process employed by Rh  ne-Poulenc for production of vitamin B₁₂ has not been revealed. However, from a variety of sources (83,86) it can be inferred that a *Pseudomonas dentrificans* producing over 200 mg L is employed. The high production is the result of classical mutation as well as (possibly) genetic engineering.

A fermentation such as that of *Pseudomonas dentrificans* typically requires 3–6 days. A submerged culture is employed with glucose, cornsteep liquor and or yeast extract, and a cobalt source (nitrate or chloride). Other minerals may be required for optimal growth. pH control at 6–7 is usually required and is achieved by ammonium or calcium salts. Under most conditions, adequate 5,6-dimethylbenzimidazole is produced in the fermentation. However, in some circumstances, supplementation may be required.

The fermentation product, which is primarily adenosylcobalamin, is retained to the largest extent within the cell. Centrifugation of the broth yields a sludge which, when dispersed in a minimum of water, water–alcohol, or water–acetone and heated, releases the vitamin into the solution. Addition of cyanide converts the cobalamins into cyanocobalamin which is then extracted from the filtered solution. Many procedures have been reported for this step, including adsorption on charcoal (87), bentonite (88), ion-exchange resins (89–91), or aluminum oxide (92). For elution, water, water–alcohol, organic bases, or hydrochloric acid are used. Chromatography on aluminum oxide and crystallization from methanol–acetone or water–acetone complete the process (93,94).

Market Forms. Vitamin B₁₂ is sold almost exclusively as cyanocobalamin. Approximately one-third of the material is for the human pharmaceutical market whereas two-thirds is used in the animal feed market, primarily for poultry and swine (see FEEDS AND FEED ADDITIVES). Modest growth in both markets has occurred in the period 1980–1995 and this trend is expected to continue.

In the human market, oral and parenteral dosage forms are prepared from the crystal. However, because of the extremely high potency, more dilute (0.1–10⁹ %) forms are available. These include dilutions with mannitol, triturations on dicalcium phosphate or resins, and spray-dried forms. Prices for these forms are driven by that of the crystal, which in early 1996 was ca \$9.50 gram (95). Prices for the vitamin have risen during the first half of the 1990s. However, little growth in price beyond inflation is anticipated.

For animal feed use, vitamin B₁₂ is usually provided in a diluted form on a carrier such as calcium carbonate and or rice hulls. An earlier practice of using a spray-dried fermentation biomass in this application appears to be no longer used.

BIBLIOGRAPHY

"Vitamin B₁₂" in *ECT* 1st ed., Vol. 14, pp. 813–828, by F. M. Robinson, Merck & Co., Inc.; "Vitamins (Vitamin B₁₂)" in *ECT* 2nd ed., Vol. 21, pp. 542–549, by S. B. Greenbaum, Diamond Shamrock Chemical Co.; "Vitamins (Vitamin B₁₂)" in *ECT* 3rd ed., Vol. 24, pp. 158–185, by W. Friedrich, Universitat Hamburg.

1. B. K  utler, in O. Isler, G. Brubacher, S. Ghisla, and B. K  utler, eds., *Vitamin II: Wasserl  sliche Vitamine*, Georg Thieme Verlag, New York, 1988, pp. 340–388.
2. L. Ellenbogen and B. A. Cooper, in L. J. Machlin, ed., *Handbook of Vitamins*, 2nd ed., Marcel Dekker, Inc., New York, 1991, pp. 491–536.
3. J. S. Combe, *Trans. Med.-Chirurg. Soc (Edinburg)*, **1**, 194 (1824).
4. T. Addison, *On the Constitutional and Local Effects of Disease of the Suprarenal Capsules*, Samuel Highley, London, 1855, p. 2.
5. G. R. Minot and W. P. Murphy, *JAMA* **91**, 923 (1926).
6. G. H. Whipple, C. W. Hooper, and F. S. Robscheit, *Am. J. Physiol.* **53**, 236 (1920).
7. W. B. Castle, *Am. J. Med. Sci.* **178**, 748 (1929).
8. M. S. Shorb and G. M. Briggs, *J. Biol. Chem.* **176**, 1463 (1948).
9. E. L. Rickes, N. G. Brink, F. R. Koniuszy, T. R. Wood, and K. Folkers, *Science* **107**, 396 (1948).
10. E. L. Smith, *Nature (London)* **161**, 638 (1948).
11. R. West, *Science* **107**, 398 (1948).
12. J. V. Pierce, A. C. Page, E. L. R. Stokstad, and T. H. Jukes, *J. Am. Chem. Soc.* **71**, 2952 (1949).
13. "The Nomenclature of Corrinoids (1973 Recommendations)," *Biochemistry* **13**, 1555 (1974).
14. D. C. Hodgkin and co-workers, *Nature (London)* **176**, 325 (1995).
15. P. G. Lenhert and D. C. Hodgkin, *Nature (London)* **102**, 937 (1961).
16. D. C. Hodgkin, *Science*, **150**, 979 (1965).
17. J. W. G. Porter, in H. C. Heinrich, ed., *Vitamin B₁₂ and Intrinsic Factor 1*, Europ  isches Symposion, Hamburg, 1956, F. Enke, Stuttgart, 1957, p. 43.
18. M. E. Coates and S. K. Kon, in Ref. 17, p. 72.
19. H. C. Heinrich, W. Friedrich, E. Gabbe, S. P. Manjrekar, and M. Staak, in *Abstracts of the 5th International Congress on Nutrition*, Washington, D.C., 1960, p. 62.
20. E. L. Smith, *Vitamin B₁₂*, Methuen, London, 1965.
21. L. Fries, *Physiol. Plantarum* **15**, 566 (1962).
22. G. G. Laties and C. Hoelle, *Phytochemistry* **6**, 49 (1967).
23. A. P. Petrosyan, L. A. Abramyan, and M. B. Sarkisyan, *Vopr. Mikrobiol.* **4**, 181 (1969).
24. J. M. Poston, *Science* **195**, 301 (1977).
25. L. J. Bogert, G. M. Briggs, and D. H. Calloway, *Nutrition and Physical Fitness*, W. B. Saunders, Philadelphia, Pa., 1973.
26. W. Friedrich, *Vitamin B₁₂ und Verwandte Corrinoid*e, Georg Thiem, Stuttgart, Germany, 1975, p. 170.
27. J. Marks, *A Guide to the Vitamins, Their Role in Health and Disease*, Medical and Technical Publishing Co., Lancaster, U.K., 1975, p. 118.
28. J. P. Glusker, in G. Litwack, ed., *Vitamins and Hormones*, Vol. 50, Academic Press, Inc., New York, 1995, pp. 1–76.
29. Y. Zhao, P. Such, and J. R  itey, *Angew. Chem. Int. Ed. Engl.* **31**, 215 (1992); M. He and P. Dowd, *J. Am. Chem. Soc.* **118**, 711 (1996).
30. T. C. Stadtman, *Science* **102**, 859 (1971).
31. R. G. Mathews, in R. L. Blakely and S. J. Benkovic, eds., *Folates and Pteridins*, Vol. 1, John Wiley & Sons, Inc., New York, 1984, p. 497.
32. C. S. Berwanger, J. Y. Jeremy, and G. Stansby, *Brit. J. Surgery* **82**, 726 (1995).
33. V. Herbert and N. Coleman, in M. E. Shils and V. R. Young, eds., *Modern Nutrition in Health and Disease*, 7th ed., Lea and Febiger, Philadelphia, Pa., 1988, pp. 388–416.
34. Food and Nutrition Board, National Research Council, *Recommended Dietary Allowances*, 10th ed., National Academy Press, Washington, D.C., 1989.
35. *The Merck Index*, 11th ed., Merck and Co., Rahway, N.J., 1989, pp. 9921–9922.
36. H. P. C. Hogenkamp, *Am. J. Clin. Nutr.* **33**, 1 (1980).
37. H. P. C. Hogenkamp, in B. M. Babior, ed., *Cobalamin-Biochemistry and Pathophysiology*, John Wiley & Sons, Inc., New York, 1975, p. 21.
38. B. M. Babior and J. S. Krouwer, *CRC Crit. Rev. Biochem.* **6**, 35 (1979).
39. R. Bonnett, in D. Dolphin, ed., *B₁₂*, John Wiley & Sons, Inc., New York, 1982, p. 201; J. Halpern, *ibid.*, p. 501; B. T. Golding, *ibid.*, p. 543.
40. E. L. Smith, *Vitamin B₁₂*, Methuen, London, 1965.
41. G. N. Schrauzer, *Acc. Chem. Res.* **1**, 97 (1968).

42. D. Dodd and M. D. Johnson, *J. Organomet. Chem.* **52**, 1 (1973).

43. G. Rytz, L. Walder, and R. Scheffold, in B. Zagalak and W. Friedrich, eds., *Vitamin B₁₂, Proceedings of the Third European Symposium on Vitamin B₁₂, Zurich, Switzerland, 1979*, Walter de Gruyter, Berlin, 1979, p. 173.

44. R. Scheffold, M. Dike, S. Dike, T. Herold, and L. Walder, *J. Am. Chem. Soc.* **102**, 3642 (1980).

45. A. Fischli, *Helv. Chim. Acta* **61**, 2560, 3028 (1978).

46. A. Fischli, *Helv. Chim. Acta* **62**, 882 (1979).

47. A. Fischli and D. Süss, *Helv. Chim. Acta* **62**, 48, 2361 (1979).

48. A. Fischli and J. J. Daly, *Helv. Chim. Acta* **63**, 1628 (1980).

49. A. Fischli and P. M. Möller, *Helv. Chim. Acta* **63**, 529, 1619 (1980).

50. Food and Nutrition Board, National Research Council, *Food Chemicals Codex*, 3rd ed., National Academy Press, Washington, D.C., 1981, p. 343.

51. *The United States Pharmacopeia XXIII(USP XXIII–NF XVIII)*, United States Pharmacopeial Convention, Inc., Rockville, Md., 1995, pp. 1719–1721.

52. H. B. Chin, J. Augustin, B. P. Klein, D. A. Becker, and P. B. Venugopal, eds., *Methods of Vitamin Assay*, 4th ed., John Wiley & Sons, Inc., New York, 1985, Chapt. 19.

53. H. L. Rosenthal, in W. H. Sebrell, Jr. and R. S. Harris, eds., *The Vitamins*, 2nd ed., Vol. II, Academic Press, Inc., New York, 1968, pp. 145–170.

54. H. R. Skeggs, in P. Gyngy and W. N. Pearson, eds., *The Vitamins*, 2nd ed., Vol. VII, Academic Press, Inc., New York, 1968, pp. 277–293; H. Baker and O. Frank *ibid.*, pp. 293–301.

55. K. Helrich, ed., *Official Methods of Analysis of the Association of Official Analytical Chemists*, 15th ed., Association of Official Analytical Chemists, Inc., Arlington, Va., 1990, pp. 952.20, 986.23.

56. Analytical Methods Committee, *Analyst* **81**, 132 (1956).

57. H. L. Newmark, J. Scheiner, M. Marcus, and M. Prabhudesai, *Am. J. Clin. Nutr.* **29**, 645 (1976).

58. P. J. Casey, K. R. Speckman, F. J. Ebert, and W. E. Hobbs, *J. Assoc. Off. Anal. Chem.* **65**, 85 (1982).

59. J. Lindemans and J. Abels, in A. P. De Leenheer, W. E. Lambert, and M. G. M. De Ruyter, eds., *Modern Chromatographic Analysis of the Vitamins*, Marcel Dekker, New York, 1985, Chapt. 12.

60. T. S. Hudson, S. Subramanian, and R. J. Allen, *J. Assoc. Off. Anal. Chem.* **67**, 996 (1984).

61. R. B. Woodward, *Pure Appl. Chem.* **17**, 519 (1968).

62. *Ibid.*, **25**, 283 (1971).

63. *Ibid.*, **33**, 145 (1973).

64. A. Eschenmoser, *Q. Rev. Chem. Soc. London* **24**, 366 (1970).

65. A. Eschenmoser, *Pure Appl. Chem. Suppl.* **2**, 69 (1971).

66. A. Eschenmoser, *Naturwissenschaften* **61**, 513 (1974).

67. A. Eschenmoser, *Chem. Soc. Rev.* **5**, 377 (1976).

68. A. Eschenmoser and C. E. Wintner, *Science* **196**, 1410 (1977).

69. A. Eschenmoser, in Ref. 43, p. 89.

70. J. H. Kreiger, *Chem. Eng. News*, 16 (1973).

71. T. H. Maugh, *Science* **179**, 266 (1973).

72. A. H. Jackson and K. M. Smith, in J. ApSimon, ed., *The Total Synthesis of Natural Products*, Vol. 1, John Wiley & Sons, Inc., New York, 1983, p. 232; *Ibid.*, Vol. 6, 1984, p. 258.

73. N. Anand, J. Bindra, and S. Ranganathan, *Art in Organic Synthesis*, 2nd ed., John Wiley & Sons, Inc., New York, 1988, p. 375.

74. R. B. Woodward, *Chem. Soc. Spec. Publ.* **21**, 217 (1967).

75. W. Friedrich, G. Gross, K. Bernhauer, and P. Zeller, *Helv. Chim. Acta* **43**, 704 (1960).

76. E. L. Smith, L. Mervyn, A. W. Johnson, and N. Shaw, *Nature (London)* **194**, 1175 (1962).

77. A. W. Johnson, L. Mervyn, N. Shaw, and E. L. Smith, *J. Chem. Soc.*, 4146 (1963).

78. K. Bernhauer and O. Möller, *Biochem., Z.* **336**, 102 (1962).

79. D. Autissier, P. Barthelemy, and L. Penasse, *Bull. Soc. Chim. Fr. Pt. II*, 192 (1980).

80. A. R. Battersby, *Acc. Chem. Res.* **26**, 15 (1993).

81. A. R. Battersby, *Pure Appl. Chem.* **65**, 1113 (1993).

82. A. R. Battersby, *Science* **264**, 1551 (1994).

83. F. Blanche and co-workers, *Angew. Chem. Int. Ed. Engl.* **34**, 383 (1995).

84. A. I. Scott, *Angew. Chem. Int. Ed. Eng.* **32**, 1223 (1993).

85. A. I. Scott, *Tetrahedron* **50**, 13315 (1994).

86. J. Crouzet, B. Cameron, F. Blanche, D. Thibaut, L. Debussche, in R. H. Baltz, G. D. Hegeman, and P. L. Skatrud, eds., *Industrial Microorganisms: Basic and Applied Molecular Genetics*, American Society for Microbiology, Washington, D.C., 1993, p. 195.

87. **U.S. Pat. 2,505,053** (Apr. 25, 1950), F. A. Kuehl and L. Chalet (to Merck & Co., Inc.).

88. **U.S. Pat. 2,626,888** (Jan. 27, 1953), S. Kutosh, G. B. Hughey, and R. Malcolmson (to Merck & Co., Inc.).

89. **U.S. Pat. 2,628,186** (Feb. 10, 1953), W. Shive (to Research Corp.).

90. **Ger. Pat. 953,643** (June 14, 1952), H. M. Shafer and A. J. Holland (to Merck & Co., Inc.).

91. H. Vogelmann and F. Wagner, *J. Chromatogr.* **76**, 359 (1973).

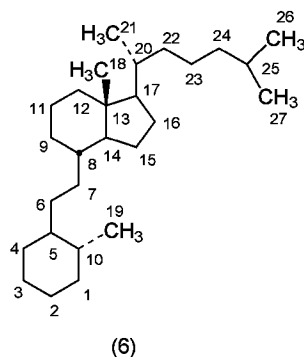
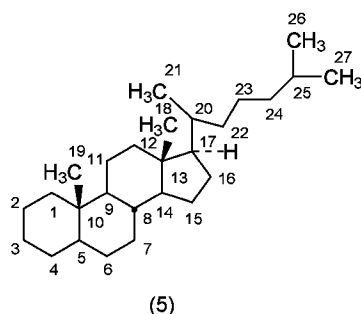
92. **Ger. Pat. 1,037,066** (Feb. 12, 1959), K. Bernhauer and W. Friedrich (to Aschaffenburg Zellstoffwerke).

93. **U.S. Pat. 2,563,794** (Aug. 7, 1951), E. L. Rickes and T. R. Wood (to Merck & Co., Inc.).

94. **U.S. Pat. 2,582,589** (Jan. 15, 1952), H. H. Fricke (to Abbott Laboratories).

95. *Chem. Mark. Rep.* (Feb. 5, 1996).

John W. Scott
Hoffmann-La Roche Inc.



In 1981, the IUPAC-IUB Joint Commission on Biochemical Nomenclature proposed that there be a set of trivial names for the important vitamin D compounds, including calciol [67-97-0] for vitamin D, calcidiol [19356-17-3] for 25-hydroxy-vitamin D₃, and calcitriol [32222-06-3] for 1 α,25-dihydroxy-vitamin D₃. This nomenclature has met with varying degrees of acceptance, as has the proposal to use calcine [69662-75-5] (deoxy-vitamin D₂) and ercalcine [68323-40-0] (deoxy-vitamin D₃) to name the triene hydrocarbon structure for 9,10-*sec*-cholesta-5,7,10(19)-triene and 9,10-*sec*-ergosta-5,7,10(19),22-tetraene, respectively. In systematic nomenclature, calcitane can be used for the basic 27-carbon skeleton instead of 9,10-secosteroid.

Isolation and Structure Determination. The isolation and structure elucidation of vitamin D are closely related to the efforts to understand and cure rickets and related bone diseases. The advent of the use of soft coal, the migration of people to cities, and the tendency of people and animals to spend less time in sunshine caused a decline in the ability of populations to synthesize sufficient quantities of vitamin D₃. This led to the increased incidence of rickets, beginning around the mid-1600s (3).

The bone disease rickets in children, and a similar condition in adult known as osteomalacia, is characterized by the body's inability to calcify the collagen matrix of growing bone, resulting in wide epiphyseal plates and large areas of uncalcified bone called osteoid. The resultant lack of rigidity of bones leads to the ends becoming twisted and bent, particularly in long bones. The ribs develop a bumpy and uneven texture known as rosary ribs and the legs become bowed. Also, the cranium becomes soft and misshapen. In adults, no long-bone growth occurs, but new bone, which is being continually remodeled, activates cells to resorb bone, followed by osteoblast-mediated bone-growth replacement (4,5). Metabolites of vitamin D act in the body to modulate these activities and maintain strong bone structure.

Although there is evidence that rickets was manifest in humans as early as 800 BC, it was not until 1645 that it was first described (6). The progress of finding a cure for rickets was relatively slow until the late 1800s, when a sufficient number of scientific developments began to allow workers to unravel the difficult puzzle of vitamin D metabolism. Tarret in 1889 had isolated ergosterol from ergot of rye and demonstrated that it was different from cholesterol. The similarity of the structures, however, led to difficulty in the structural elucidation of the vitamin D molecule. Mellanby (7) demonstrated the lack of a dietary component could be used to develop rickets and was able to raise D-deficient animals. This allowed research aimed at finding the antirachitic factor to be carried out.

In 1919, Huldschinski (8) realized that uv light cured rickets and impacted on its etiology. The uv light and cod liver oil were found to be useful in the treatment of the disease, and irradiation of food produced the same effect as irradiation of the animal. The link between irradiation and plant materials led to the conclusion that ergosterol was an antirachitic substance, and an extensive effort was made to characterize the chemistry of irradiated ergosterol.

Windaus in 1933 derived the skeletal structure of vitamin D₂. He found the elemental formula to be C₂₈H₄₄O. The side chain was unequivocally characterized by x-ray crystallography later, in 1948. Reindel and Kipphan (9) in 1932 ozonized the side chain and elucidated the C₂₂ double bond. The reaction to form a Diels Alder adduct with maleic anhydride and the typical diene uv spectrum led to the conclusion there was a diene in the B ring. In 1934 Fernholz and Chakravorty (10) determined the provitamin had a 3 β-hydroxy group because the molecule could be precipitated with digitonin, a reaction characteristic of the 3 β-ol functionality in steroids.

Perbenzoic acid gave a doubly unsaturated triol monobenzoate. Only two hydroxyl groups could be acetylated, and one was tertiary. The saturated triol reacted with lead tetracetate to give an α glycol. When reacted with chromic acid, it gave a hydroxy lactone. From these observations, Windaus and Grundmann (11) described the correct structure for ergosterol (1).

The observation that the uv spectrum of provitamin D changed with uv irradiation and also produced antirachitic activity led to the conclusion that vitamin D was derived from the provitamin. Windaus found the vitamin D₂ formula C₂₈H₄₄O to be isomeric with the provitamins.

Bunzl in 1932 used x-ray crystallography to demonstrate that many of the preparations described in the literature were mixtures and that vitamin D₂ was one crystal structure.

Calciferol, when hydrogenated catalytically, took up 4 moles of hydrogen and gave a compound with the empirical formula C₂₈H₅₂O. Sodium in ethanol reduction gave a dihydropyridine that reacted with 3 moles of perbenzoic acid, thus demonstrating the derivative to have three double bonds.

Ozonolysis of vitamin D₂ gave 2,3-dimethylbutanol, showing the side chain to contain the 22 double bond and a C₂₄ methyl group.

Vitamin D₂ reacted with maleic anhydride to give a mono Diels-Alder adduct, which hydrolyzed to yield a dicarboxylic acid. Acetylation of the alcohols, esterification of carboxylic acids, and hydrogenation gave a compound that, when ozonized, gave a saturated ketone, C₁₉H₃₄O. This molecule contains the C and D ring and side chain from vitamin D₂, indicating that the B ring must have been opened by photolytic cleavage of the C₉-C₁₀ bond (12). The vitamin D₂ molecule was thus shown to be a tricyclic material with four double bonds.

Windaus and Boch (13) isolated and characterized 7-dehydrocholesterol in 1937 from pig skin. They further showed that vitamin D₃ could be generated from the provitamin by uv irradiation.

Irradiated ergosterol was found not to be as antirachitic in the chick as in the rat, whereas the chick could be protected by direct irradiation. The provitamin in cholesterol was shown not to be ergosterol. Rygh (14) in 1935 found that 1 rat unit of cod liver oil was 100 times more potent in chicks than 1 rat unit of vitamin D₂. Brockmann (15) in 1936, prepared the pure crystalline 3,5-dinitrobenzoate derivative of vitamin D₃ obtained from tuna liver oil

(subsequently demonstrated in other species like halibut and blue fin tuna). This material was shown to be identical to the dinitrobenzoate derivative of vitamin D₃ obtained by Windaus from the irradiation of 7-dehydrocholesterol. Thus, the chemical identity of vitamin D₂ and D₃ were well established. A more detailed description of the events leading to the structure elucidation of the vitamin Ds can be found in Reference 6.

Occurrence. The provitamins, precursors of the vitamin Ds, are distributed widely in nature, whereas the vitamins themselves are less prevalent. The amounts of provitamins D₂ and D₃ in various plants and animals are listed in Table 2.

Table 2. Occurrence of the Provitamins D in Selected Plants and Animals, Parts per Thousand of Total Sterol

Source	Amount	Source	Amount
cottonseed oil	28	gallstones, man	0.25
rye grass	15	<i>Mytilus edulis</i> , sea mussel	100
wheat germ oil	10	<i>Modiolus demissus</i> , ribbed mussel	370
carrot	1.7	<i>Ostrea virginica</i> , oyster	80
cabbage	0.5	<i>Ostrea edulis</i> , oyster	34
skin, pig	46	<i>Asterias rubens</i> , starfish	3.8
skin, chicken feet	25	common sponges	20
skin, rat	19	common coral	10
skin, mouse	9	<i>Aspergillus niger</i> , mold	1000
skin, calf	7	<i>Cortinellus shiitake</i> , mushroom	1000
skin, human infant	1.5	<i>Claviceps purpurea</i> , ergot	9000
skin, human adult	4.2	<i>Saccharomyces cerevisiae</i> , yeast	800
liver, Japanese tuna	11	<i>Penicillium puberulum</i> , mold	1000
liver, Atlantic cod	4.4	<i>Fucus vesiculosus</i> , alga, seaweed	0.8
liver, shark	1.0	<i>Tubifex</i> sp, waterworm	210
liver, halibut	0.6	<i>Bumbricus terrestris</i> , earthworm	170
liver, tuna	1.0	<i>Tenebrio molitor</i> , mealworm	120
eggs, Chinese duck	60	<i>Cyronomus</i> , sp, goat	61
eggs, cod (roe)	5.5	<i>Cancer pagurus</i> , common crab	15
eggs, hen	1.6	<i>Dephnia</i> sp, water flea	7.5
wool fat, sheep	3.9	<i>Musca domestica</i> , housefly	7.0
milk, cow	2.3	<i>Crangon vulgaris</i> , shrimp	3.8
pancreas, beef	1.8	<i>Homarus vulgaris</i> , lobster	2.5
spinal cord, beef	1.2	<i>Helix pomatia</i> , edible snail	97
blood serum, cow	0.5	<i>Arion empiricorum</i> , slug, red road snail	220
herring oil	0.5	<i>Littorina littorea</i> , periwinkle	170
heart, calf	0.32	<i>Sepia</i> sp, cuttlefish	12

Fish-liver oil, liver, milk, and eggs are good natural sources of the D₃ vitamin. Most milk sold in the United States is fortified with manufactured vitamin D. Fish oil is the only commercial source of natural vitamin D₃, and the content of the vitamin varies according to species as well as geographically, ie, Atlantic cod contain 100 IU g where IU (International Unit) = 0.025 µg of vitamin D₃, whereas oriental tuna (*Percomorpli*) contain 45,000 IU g oil.

Vitamin D₃ rarely occurs in plants. However, *Solanum glaucophyllum*, *Solanum malacocylon*, *Cestrum diurnum*, and *Trinetum flavescens* have been shown to contain water-soluble glycosides of vitamin D analogues with 1 α,25-dihydroxy-vitamin D activity (16–22). The vitamin D content in various plant and animal materials is shown in Table 3. Vitamin D₃ occurs naturally in all animals (24).

Table 3. Distribution of Vitamin D Activity^a

Sample	Amount ^b
phytoplankton	0
sargassum (a gulfweed)	some activity
clover hay	
sun-cured	slight
dark-cured	0
mushrooms (<i>Agaricus campestris</i>)	0.21 IU g
milk (unfortified)	
winter (bovine)	5.3 IU L
summer (bovine)	53 IU L
milk (human)	63 IU L
milk colostrum (human)	315–635 IU L
egg yolk	150–400 IU g
butter	4–8 IU g
fish-liver oils	50–45,000 IU g

^a Ref. 23.

^b IU (International Unit) = 0.025 µg of vitamin D₃.

Chemical and Physical Properties

Provitamin. The chemistry of the D vitamins is intimately involved with that of their precursors, the provitamins. The manufacture of the vitamins and their derivatives usually involves the synthesis of the provitamins, from which the vitamin is then generated by uv irradiation. The chemical and physical properties of the provitamins are discussed below, followed by the properties of the vitamins.

3 β-Hydroxy steroids which contain the 5,7-diene system and can be activated with uv light to produce vitamin D compounds are called provitamins. The two most important provitamins are ergosterol (1) and 7-dehydrocholesterol (3). They are produced in plants and animals, respectively, and 7-dehydrocholesterol is produced synthetically on a commercial scale. Small amounts of hydroxylated derivatives of the provitamins have been synthesized in efforts to prepare the metabolites of vitamin D, but these products do not occur naturally. The provitamins do not possess physiological activities, with the exception that provitamin D₃ is found in the skin of animals and acts as a precursor to vitamin D₃, and synthetic dihydroxalated

analogues of pro- and previtamins have been found to have selective activity towards nuclear and nonnuclear receptors (24–27).

Provitamin D₂. Ergosterol is isolated exclusively from plant sources. The commercial product is ca 90–100% pure and often contains up to 5 wt % of 5,6-dihydroergosterol. Usually, the isolation of provitamin D₂ from natural sources involves the isolation of the total sterol content, followed by the separation of the provitamin from the other sterols. The isolation of the sterol fraction involves extraction of the total fat component, its saponification, and then reextraction of the unsaponifiable portion with an ether. The sterols are in the unsaponifiable portion. Another method is the saponification of the total material, followed by isolation of the nonsaponifiable fraction. Separation of the sterols from the unsaponifiable fraction is done by crystallization from a suitable solvent, eg, acetone or alcohol. Ethylene dichloride, alone or mixed with methanol, has been used commercially for recrystallization. In the case of yeasts, it is particularly difficult to remove the ergosterol by simple extraction, thereby obtaining only ca 25% recovery. Industrially, therefore, the ergosterol is obtained by preliminary digestion with hot alkalies or with amines (28–33). Variations of the isolation procedure have been developed. For example, after saponification, the fatty acids may be precipitated as calcium salts, which tend to absorb the sterols. The latter are then recovered from the dried precipitate by solvent extraction.

Provitamin D₃. Provitamin D₃ is made from cholesterol, and its commercial production begins with the isolation of cholesterol from one of its natural sources. Cholesterol occurs in many animals, and is generally extracted from wool grease obtained by washing wool after it is sheared from sheep. This grease is a mixture of fatty-acid esters, which contain ca 15 wt % cholesterol. The alcohol fraction is obtained after saponification, and the cholesterol is separated, usually by complexation with zinc chloride, followed by decomplexation and crystallization. Cholesterol can also be extracted from the spinal cords and brains of animals, especially cattle, and from fish oils.

Cholesterol (7) is converted to 7-dehydrocholesterol (3) (see Fig. 1). This process usually involves the Ziegler allylic bromination of the 7 position followed by dehydrobromination (34). Esterification of the cholesterol (8) is necessary to prevent oxidation of the 3 β -alcohol by the brominating agent. Allylic bromination may be accomplished with a variety of brominating agents, eg, *N*-bromosuccinimide, *N*-bromophthalimide, or preferably 5,5-dimethyl-1,3-dibromohydantoin (35). Bromine in carbon disulfide can be used if the free-radical bromination is photocatalyzed (36). A mixture of 7 α - (9) and 7 β -bromo cholesteryl esters (10) is obtained and treated with an appropriate base to dehydrohalogenate the molecule and give the 7-dehydrocholesteryl ester (11) (37,38). Proper conditions for this reaction are necessary to generate a high yield of the desired 7-dehydro product instead of the undesired cholesta-4,6-dien-3 β -ol ester (12), which is formed as a by-product. Various reagents can be used to perform the dehydrohalogenation. Trimethyl phosphite or pyridine bases, particularly trimethylpyridine, have been used; the symmetrical collidine is the reagent of choice (35,38,39). *t*-Butylammonium fluoride has also been used to improve the yield of high quality 5,7-diene (40).

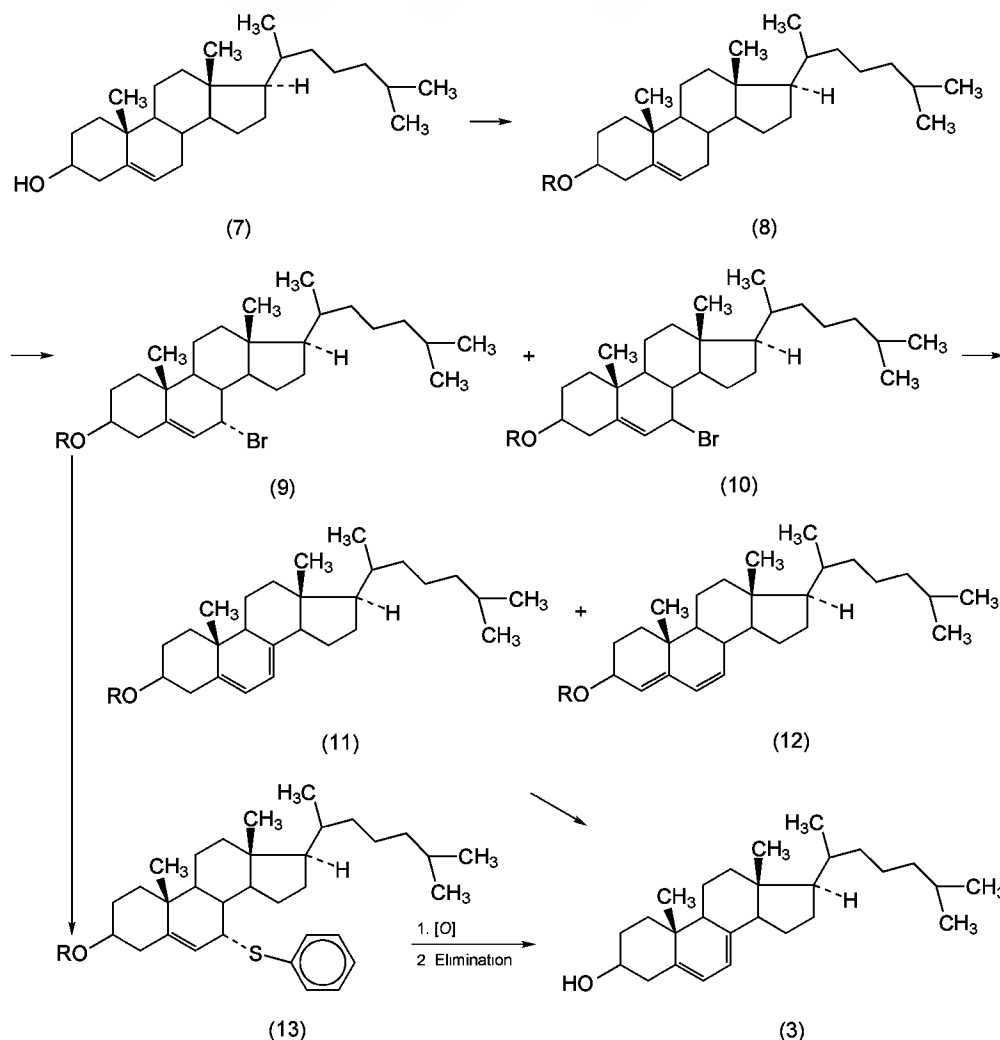


Fig. 1. Conversion of cholesterol to 7-dehydrocholesterol.

The 7 α -bromo steroid (9) can also be treated with sodium phenyl selenolate (41). The resultant 7 β -phenyl selenide (13) can be oxidized and the corresponding phenyl selenoxide eliminated to form the 7-dehydrocholesteryl ester (11).

7-Dehydrocholesterol has also been made from cholesterol by the Windaus procedure (Fig. 2); the 3,7-dibenzoate (16) is obtained (via (14) and (15) by oxidation and reduction), which undergoes thermal elimination to give the 7-dehydrocholesteryl benzoate (11) (42–44). However, the yields are substantially lower than those achieved by the bromination–dehydrobromination method.

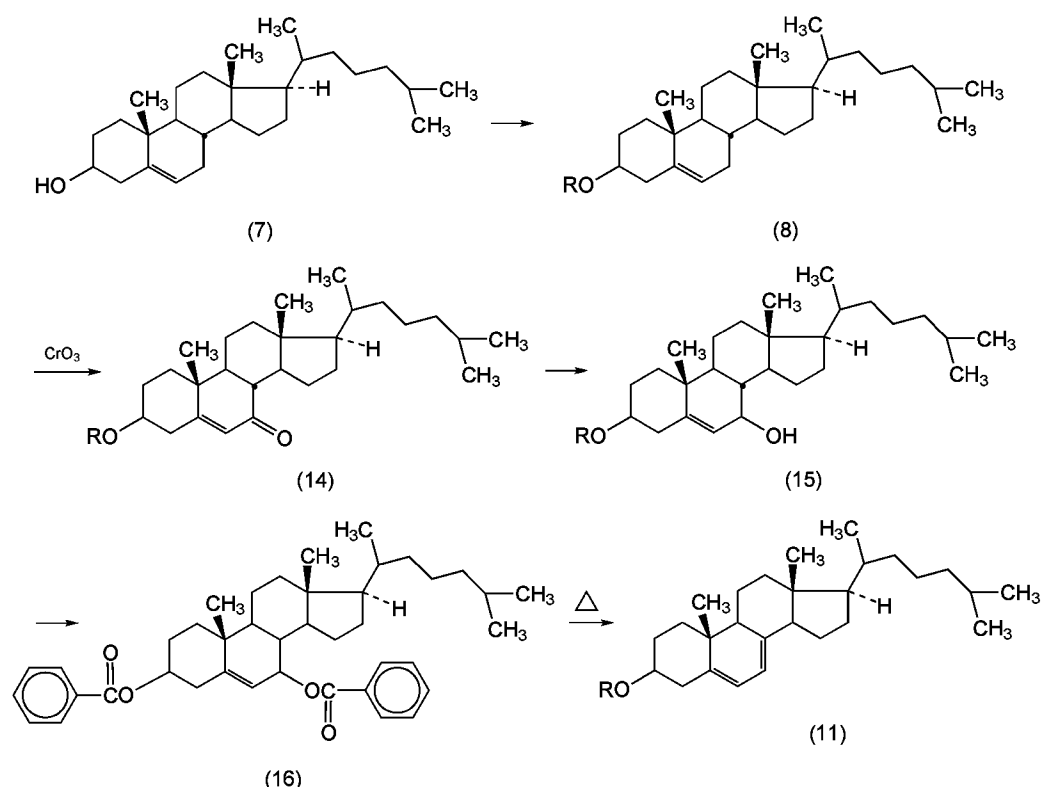


Fig. 2. The Winhaus procedure (42).

7-Tosylhydrazone and 7-phenyl sulfoxide groups have also been introduced into cholesterol and eliminated to prepare the 5,7-diene (45,46). The method of choice is the allylic bromination–dehydrobromination procedures, and the commercial yields in converting cholesterol to 7-dehydrocholesterol are in the range of 35–50%.

Vitamin D. The irradiation of the provitamins to produce vitamin D as well as several isomeric substances was first studied with ergosterol. The chemistry is identical for the vitamin D₃ series and yields analogous isomers. In 1932, a scheme for the irradiation of ergosterol leading to vitamin D₂ was proposed (45). Twenty years later, the mechanism of the irradiation of the provitamins to vitamin D and its photoisomers was further elucidated (46,47). More recently, Jacobs (48) has reviewed the photochemistry. A number of products associated with the irradiation process are shown in Figure 3. The geometry and electronic characteristics of these molecules have been well established by x-ray crystallographic analysis and valence force-field calculations (50–52). The irradiation process, which converts 7-dehydrocholesterol to vitamin D also occurs in the skin of animals if sufficient sunlight is available. The photochemical and thermal isomerizations occur during the generation of vitamin D *in vivo* as well as during its synthetic photochemical preparation (24,53,54). The initial step involves ring opening of the B-ring of the sterol by ultraviolet activation of the conjugated diene. The absorbance of uv energy activates the molecule, and the $\pi \rightarrow \pi^*$ excitation (absorption, 250–310 nm; $\lambda_{\text{max}} = 291 \text{ nm}$, $\epsilon = 12,000$) results in the opening of the 9,10 bond and the formation of the (Z)-hexadiene, previtamin D₃ (R) [1173-13-3] (17) or previtamin D₂ (R') [21307-015-1] (18). The uv irradiation of 7-dehydrocholesterol or ergosterol results in the steady diminution in concentration of the provitamin, initially giving rise to predominantly previtamin D. The pre- levels reach a maximum as the provitamin level drops below ca 10%. The concentration of the previtamin then falls as it is converted to tachysterol and lumisterol, which increase in concentration with continued irradiation (see Fig. 4). Temperature, frequency of light, time of irradiation, and concentration of substrate all affect the ratio of products. Previtamin D undergoes thermal equilibration to vitamin D (2) or (4), *cis*-vitamin D₂ [50-14-6] and *cis*-vitamin D₃ [67-97-0], respectively. The conversion of previtamin D (18) or (17) at temperatures of <80°C by thermal isomerization to give the *cis* vitamin (ergocalciferol) (2) or cholecalciferol (4) involves an equilibrium, as indicated in Table 4 (55,56).

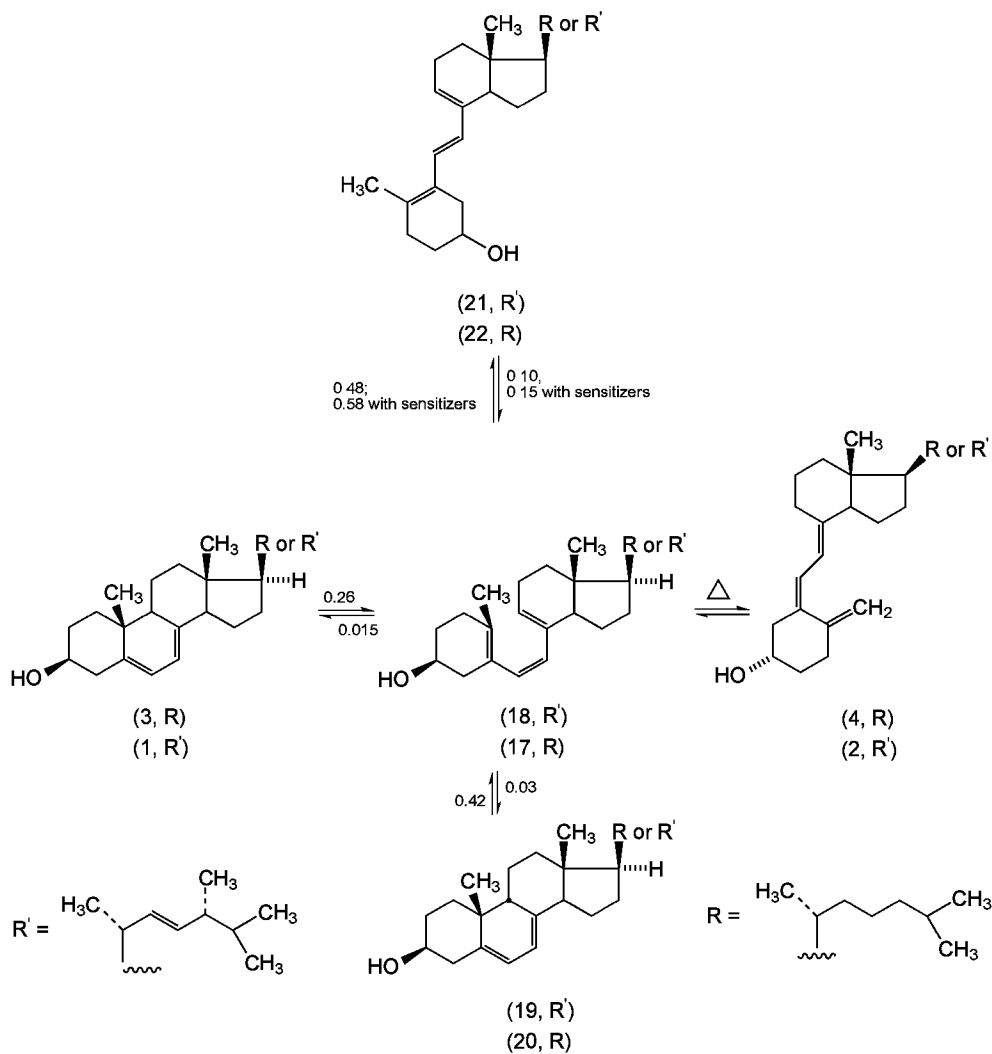


Fig. 3. Photochemical and thermal isomerization products of vitamin D manufacture (49). The quantum yields of the reactions are listed beside the arrows for the given reactions.

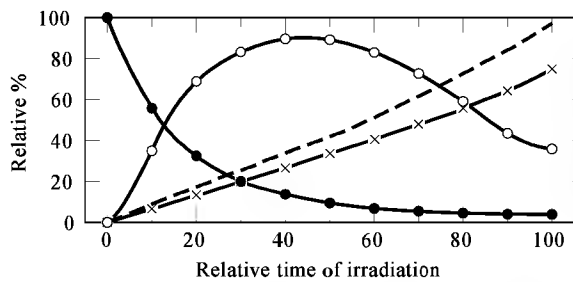


Fig. 4. Time course for uv-irradiation of 7-dehydrocholesterol (●): (○), previtamin D; (x), lumisterol; (---), tachysterol.

Table 4. Inverconversion Time of Previtamin D₃ and Vitamin D₃ at 20°C, and 40°C^a

Formation, °	Vitamin D ₃ from previtamin D ₃			Previtamin D ₃ from vitamin D ₃	
	Days at 20°C	Days at 20°C	Hours at 40°C	Days at 20°C	Hours at 40°C
2	27	0.2	0.4	2.2	3.7
5	68	0.4	1.1	8.2	11.2
7	96	0.5	1.5		18.6
10	140	0.8	2.3		43.3
20	269	1.6	4.8		
30	474	2.6	7.8		
40	681	3.8	11.3		
50	926	5.2	15.6		
60	1230	7.0	21.2		
70	1628	9.5	29.2		
80	2204	13.3	43.4		
90	2910	23.3			

^a Ref. 55.

The equilibrium composition is normally ca 80° vitamin D and 20° previtamin D. This reaction is an intramolecular {1–7} H sigmatropic shift

and occurs through a rigid cyclic transition state (57).

Additionally, the $\pi \rightarrow \pi^*$ excitation of previtamin D can result in ring closure back to the provitamin (1) [57-87-4] or (3) [434-16-2] or to lumisterol₂ [474-69-1] (19) or lumisterol₃ [5226-01-7] (20), which have the 9 β ,10 α configuration. This excitation can also exhibit (*Z*) $\leftrightarrow$ (*E*) photoisomerization to the 6,7-(*E*)-isomer, tachysterol₂ [115-61-7] (21) or tachysterol₃ [63902-44-3] (22) (58,52).

Other photoinduced cyclization reactions can occur by conrotatory bond formation to give the 9 β ,10 β -antiisomers, isopyrocalciferol₂ [474-70-4] (23) or isopyrocalciferol₃ [10346-44-8] (24) (Fig. 5), whereas thermal cyclization at $>100^\circ\text{C}$ leads to the two 9,10-syn isomers, (9 α ,10 α)-pyrocalciferol (27) [128-27-8] or (28) [10346-43-7] by a disrotatory bond formation mechanism (47). Ultraviolet over-irradiation leads to photopyro- (29) [41411-05-6] or (30) and photoisopyrocalciferols (25) [26241-65-6] or (26) [85354-28-5], respectively and to the formation of suprasterols of the type shown in structure (37) (59,60) (Fig. 6). Prolonged irradiation of the mixture of isomers can also lead to toxisterols of the type shown in structures (38) and (39), where R = D₂ and R' = D₃ side chains (see Fig. 3) (61–63). More than 20 members of this type of substance have been identified. There is little evidence that these materials are toxic; however, their nomenclature leads to some misunderstanding. These compounds generally show little if any biological activity and have not been found *in vivo* (15,63–66). Normal irradiation conditions for the production of vitamin D include sufficiently low temperatures so that these products do not form, and they are not found in normal commercial samples of D₃ resins.

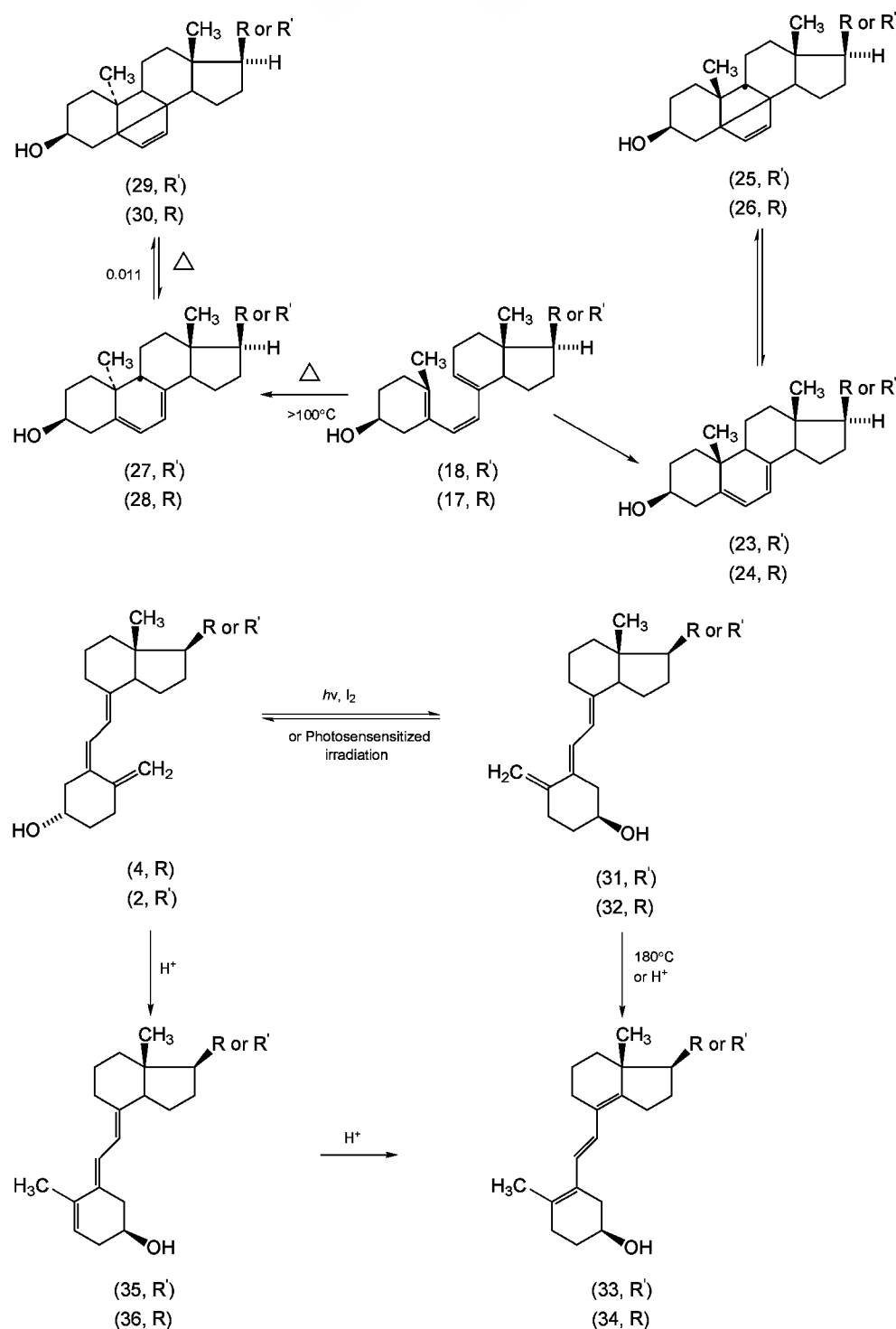


Fig. 5. Products of the isomerization of pre- and *cis*-vitamin D.

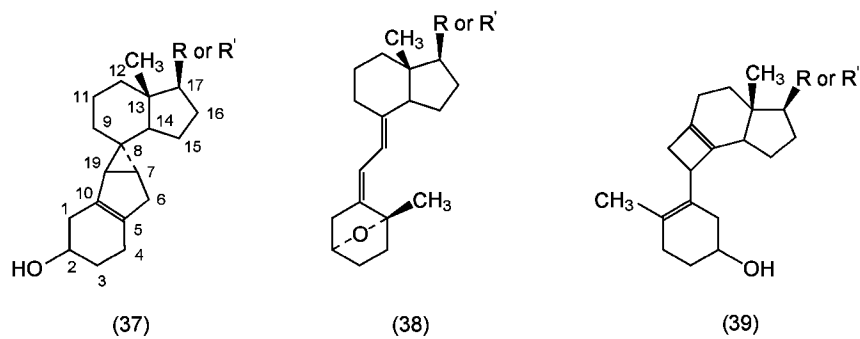


Fig. 6. Products of overirradiation of vitamin D: suprasterol II (37), toxisterol-E (38), and toxisterol E₁ (39).

The irradiation of calciferol in the presence of iodine leads to the formation of 5,6-*trans*-vitamin D₂ [14449-19-5] (31) or D₃ [22350-41-0] (32) (67,68). 5,6-*trans*-Vitamin D as well as vitamin D (2) or (4) can be converted to isovitamin D by treatment with mineral or Lewis acids. Isocalciferol (35) [469-05-6] or (36) [42607-12-5] also forms upon heating of 5,6-*trans*-vitamin D. Isotachysterol (33) [469-06-7] or (34) [22350-43-2] forms from isocalciferol or vitamin D upon treatment with acid, and its production appears to be the result of sequential formation of *trans*- and isocalciferol from calciferol. These reactions are the basis of the antimony trichloride test for vitamin D (69–72).

Commercially, the irradiation of the 5,7-diene provitamin to make vitamin D must be performed under conditions that optimize the production of the provitamin while avoiding the development of the unwated isomers. The optimization is achieved by controlling the extent of irradiation, as well as the wavelength of the light source. The best frequency for the irradiation to form previtamin is 295 nm (64–66). The unwanted conversion of previtamin to tachysterol is favored when 254 nm light is used. Sensitized irradiation, eg, with fluorenone, has been used to favor the reverse, triplet-state conversion of tachysterol to previtamin D (73,74).

The molecular extinction coefficients (at various wavelengths) of the four main components of the irradiation are shown in Table 5. The absorption of light above 300 nm is favored by tachysterol. A yield of 83% of the previtamin at 95% conversion of 7-dehydrocholesterol can be obtained by irradiation first at 254 nm, followed by reirradiation at 350 nm with a yttrium aluminum garnet (YAG) laser to convert tachysterol to previtamin D. A similar approach with laser irradiation at 248 nm (KrF) and 337 nm (N₂) has also been described (76).

Table 5. Molecular Extinction Coefficients of Irradiation Products^a

λ, nm	7-Dehydrocholesterol	Previtamin D ₃	Tachysterol ₃	Lumisterol ₃
254	4,500	725	11,450	4,130
300	1,250	930	11,250	1,320
330	25	105	2,940	30
340	20	40	242	25
350	10	25	100	20

^a Ref. 75.

The irradiation of the provitamin has been achieved using the acetate and benzoate esters, although the free alcohol form of the provitamin is usually used (77).

Physical Properties. The physical properties of the provitamins and vitamins D₂ and D₃ are listed in Table 6. The values are listed for the pure substances. The D vitamins are fat-soluble and, as such, are hydrophobic.

Table 6. Physical Properties of Provitamins and Vitamins D₂ and D₃

Properties	Substance			
	7-Dehydro-cholesterol	Ergosterol	Vitamin D ₂ (ergocalciferol)	Vitamin D ₃ (cholecalciferol)
melting point, °C	150–151	165	115–118	84–85
color and form	solvated plates from ether–methanol	hydrated plates from alcohol; needles from acetone	colorless prisms from acetone	fine colorless needles from dilute acetone
CAS Registry Number	[434-16-2]	[57-87-4]	[50-14-6]	[67-97-0]
optical rotation (α ²⁰ _D), °				
acetone			82.6	83.3
ethanol			103	105–112
chloroform	113.6	135	52	51.9
ether			91.2	
petroleum ether			33.3	
benzene	127.1			
coefficient of rotation per °C in alcohol			0.515	
uv max, nm	282	281.5	264.5	264.5
specific absorption, E _{max} , at 1%	308		458.9 ± 7.5	473.2 ± 7.8
conc				
potency ^a , IU/g			40 × 10 ⁶	40 × 10 ⁶
biological activity			in mammals	in mammals and birds
chicken efficacy, %			8–10 ^b	100
solubility, g/100 mL				
acetone at 7°C			7	
acetone at 26°C			25	
absolute ethanol	sl sol		28	

at 26°C	0.15			
at Δ°C	2.2			
ethyl acetate at 26°C			31	
water	insol	insol	insol	insol

^a The international standard for vitamin D is an oil solution of activated 7-dehydrocholesterol (3). The IU is the biological activity of 0.025 µg of pure cholecalciferol.

^b Studies have claimed an efficacy as high as 10⁰ o (77).

Shipping and Handling. Vitamin D and its products are sensitive to uv light, heat, air, and mineral acids. Its sensitivity to these conditions is exaggerated by the presence of heavy-metal ions, eg, iron. Care should be taken to store and ship vitamin D and its various product forms so that exposure to these conditions is minimized. Pharmaceutical grade vitamin D₃ is supplied in the pure crystalline form. This is used in vitamin D as well as multivitamin preparations. Commercial sources of feed-grade vitamin D are usually the vitamin D₃ resin stabilized by spray- or roll drying a starch or gelatin suspension of the vitamin. These products should be stored in a cool, dry place in opaque, hermetically sealed containers under nitrogen. Vitamin D is generally recognized as safe when used in accordance with good manufacturing or feeding practices (79).

Shipping vitamin D in crystalline or resin form should be done in containers marked appropriately to indicate the material is toxic by DOT standards. Its proper DOT labeling is DOT Hazard Class 6.1, poisonous. Waste material should be burned or placed in an appropriate landfill.

The provitamins are also unstable to heat and light. They generally should be stored in a dark, cool, dry place. The provitamin is more stable if shipped with 10–15 wt % methanol rather than in a dry form.

Analytical and Test Methods

The development of reliable uv analysis permitted the dependable detection and assay of the provitamins and vitamins. Prior to this, the Lieberman-Bouchard chemical test was used, but the color reaction gave many false positives and was relatively inaccurate.

In 1949 the World Health Organization adopted the biological activity of 1 mg of an oil solution containing 0.025 µg of crystalline D₃ as the analytical standard for vitamin D₃. This standard was discontinued in 1972. USP uses crystalline cholecalciferol as a standard (80). Samples of reference standard may be purchased from U.S. Pharmacopeial Convention, Inc., Reference Standards Order Department, 12601, Twinbrook Parkway, Rockville, Maryland 20852. One international unit of vitamin D activity is that activity demonstrated by 0.025 µ of pure crystalline *cis*-vitamin D₃. One gram of vitamin D₃ is equivalent to 40 × 10⁶ IU or USP units. The international chick unit (ICU) is identical to the USP unit.

USP also issues vitamin D₃ capsules for AOAC determination in rats and an oil solution for the vitamin D₃ AOAC determination in chicks. Historically, the following units (shown with their approximate international unit equivalence) have been used but are currently abandoned: 1 clinical unit 12–17 IU; 1 biological unit 0.125 IU; 1 protection unit 0.125 IU; 1 Laquer unit 0.14 IU; 1 Poulson unit 0.2 IU; 1 Steenbach unit 3 IU. The MRC, ICU, and Coward units all approximated the international unit and are also no longer in common use.

The standard chemical and biological methods of analysis are those accepted by the *United States Pharmacopeia XXIII* as well as the ones accepted by the AOAC in 1995 (81–84). The USP method involves saponification of the sample (dry concentrate, premix, powder, capsule, tablet, or aqueous suspension) with aqueous alcoholic KOH; solvent extraction; solvent removal; chromatographic separation of vitamin D from extraneous ingredients; and colormetric determination with antimony trichloride and comparison with a solution of USP cholecalciferol reference standard.

The AOAC (978.42) recognizes a similar procedure, except that the unsaponifiable material is treated with maleic anhydride to remove the trans-isomer which may possibly be present (83). The antimony trichloride colorimetric assay is performed on the trans-isomer-free material. This procedure cannot be used to distinguish certain inactive isomers, eg, isotachysterol; if present, these are included in the result, giving rise to a falsely high analysis. A test must therefore be performed to check for the presence of isotachysterol.

Preferably, high pressure liquid chromatography (hplc) is used to separate the active pre- and cis-isomers of vitamin D₃ from other isomers and allows their analysis by comparison with the chromatograph of a sample of pure reference *cis*-vitamin D₃, which is equilibrated to a mixture of pre- and cis-isomers (82,84,85). This method is more sensitive and provides information on isomer distribution as well as the active pre- and cis-isomer content of a vitamin D sample. It is applicable to most forms of vitamin D, including the more dilute formulations, ie, multivitamin preparations containing at least 1 IU/g (AOAC Methods 979.24; 980.26; 981.17; 982.29; 985.27) (82). The practical problem of isolation of the vitamin material from interfering and extraneous components is the limiting factor in the assay of low level formulations.

A number of methods have been developed for the chromatographic separation of vitamin D and related substances and can be found in Table 7.

Table 7. Methods for Chromatographic Separation of Vitamin D and Related Steroid^a

Method	Comments
<i>Paper chromatography</i>	
quinoline-impregnated paper	
reversed-phase paper	
column chromatography	
alumina	
floridin	
celite	useful for multivitamin tablets
silicic acid	useful for metabolites
factice	can resolve vitamin D ₂ and vitamin D ₃
Sephadex LH-20	highly useful for metabolites
<i>High pressure liquid chromatography</i>	
Zorbax-Sil support	separates 24(R), 25-dihydroxy-vitamin D ₃ from 24(S), 25-dihydroxy-vitamin D ₃
	useful for metabolites
	assay of 25-hydroxy-vitamin D ₃
ODS ^b -Permaphase	useful for vitamin D metabolites
silica gel	commercial vitamin D ₃ assay
<i>Thin-layer chromatography</i>	
silicic acid	effective for irradiation mixtures
silica gel G	two-dimensional
<i>Gas chromatography</i>	
	separates trimethylsilyl ethers of vitamin D ₃
	cyclizes vitamin D ₃
	useful for 25-hydroxy-vitamin D ₃
	useful for multivitamin tablets

useful for vitamin D₂ in milk

^a Ref. 86.
^b ODS octadecyl (C₁₈) silane.

Biological Assay. The USP and AOAC recognize a biological method for the determination of vitamin D. The rat line test, however, is slow, expensive and not as accurate as the chemical or chromatographic methods. Rachitic rats are fed diets containing the vitamin D sample. This test measures bone growth on the proximal end of the tibia or distal end of the ulna, which is visualized by staining with silver nitrate. This test is not applicable to products offered for poultry feeding. The AOAC recognizes another procedure which measures the vitamin D sample activity in increasing bone ash of growing chicks compared to the activity of a USP cholecalciferol reference standard. It too is slow, expensive, and gives variable results.

Physical Methods. Vitamins D₂ and D₃ exhibit uv absorption curves that have a maximum at 264 nm and an E_{max} (absorbance) of 450–490 at 1° concentration (Table 8). The various isomers of vitamin D exhibit characteristically different uv absorption curves. Mixtures of the isomers are difficult to distinguish. However, when chromatographically separated by hplc, the peaks can be identified by stop-flow techniques based on uv absorption scanning or by photodiodearray spectroscopy. The combination of elution time and characteristic uv absorption curves can be used to identify the isomers present in a sample of vitamin D.

Table 8. Approximate uv Absorbance of 7-Dehydrocholesterol, Pre- and *cis*-Vitamin D

Absorbance at μg	(E 1° o 1 cm)		
	7-Dehydrocholesterol	Previtamin D	<i>cis</i> -Vitamin D
230	35	190	250
235	40	170	280
240	50	210	325
245	60	235	360
250	85	250	400
255	120	260	430
260	150	270	460
265	200	265	470
270	270	250	450
273	282		
275	250	220	420
277	240		
280	290	180	340
282	293		
285	240	150	300
290	155	100	180
295	170	70	120
300	70	50	80

Infrared and nmr spectroscopy have been used to help distinguish between vitamins D₂ and D₃ (87–89). X-ray crystallographic techniques are used to determine the vitamin D structure, and gas chromatography also is a method for assaying vitamin D (49,90–95).

Provitamin D. The molecular extinction coefficient of 7-dehydrocholesterol at 282 nm is 11,300 and is used as a measure of 7-dehydro isomer content of the provitamin (96,97). High pressure liquid chromatography can also be used to analyze the provitamins. There are a variety of chemicals that show characteristic colors when reacted with the provitamins. Some of these are listed in Table 9.

Table 9. Chemical Test Methods for Provitamin D

Name of reaction	Components	Results	Interpretation
revised Salkowski reaction	CHCl ₃ + H ₂ SO ₄ (conc)	deep red acid layer	differentiates from sterols lacking conjugated diene (red color in CHCl ₃ layer) acid gives green fluorescence
Lieberman-Burchard reaction	CHCl ₃ ; acetic acid–H ₂ SO ₄ added dropwise	red color develops and changes to blue-violet to green	can be quantitative; acts similarly, but red color lasts longer
Tortelli-Jaffr� reaction	acetic acid + 2 wt% Br ₂ in CHCl ₃	green	sterols with ditertiary double bonds; vitamin D and compounds that give similar bonds upon isomerization or reaction
Rosenheim reaction	CHCl ₃ + trichloroacetic acid in H ₂ O	red color develops and changes to light blue	
Rosenheim reaction	CHCl ₃ + lead tetraacetate in CH ₃ COOH is added; then trichloroacetic acid is added	green fluorescence	not given by esters of provitamin D; can be used to distinguish be-tween provitamin and provitamin ester;
chloral hydrate	mixture of crystalline provita-mins and chloral hydrate heated slowly; melts at 50�C	color develops and changes red to green to deep blue	sensitive to 0.1 μg and is quantitative other sterols, eg, cholesterol, do not react to give color
antimony trichloride reaction	CHCl ₃ + SbCl ₃	red color	
Chugaev reaction	glacial acetic acid plus acetyl chloride and zinc chloride heated to boiling	eosin-red greenish yellow fluorescence	1:80,000 sensitivity

The extremely low levels of vitamin D and its metabolites in biological systems make it very difficult to assay these products by traditional methods. Calcium-binding protein is not found in the intestinal mucosa of vitamin D-deficient animals. It is synthesized only in response to the presence of a material with vitamin D activity. Thus, using antiserum specific to intestinal calcium-binding protein, a radioimmunoassay (98) conducted on homogenates of intestinal mucosa of chicks fed the test material for 7–10 d allows assay of the material for vitamin D activity down to 1–200 IU. The

development of this technique has allowed research to be conducted since the mid-1970s to elucidate the vitamin D metabolism and biochemistry and is now used routinely to assay vitamin D-active materials in clinical as well as research samples (6,40,51,53,55).

Synthesis

Manufacture. Most of the vitamin D produced in the world is made by the photochemical conversion of 7-dehydrocholesterol. Ergosterol is not used as extensively as it once was, because it offers no real price advantage and, upon irradiation, it gives vitamin D₂, which has been shown to be less active in many species. The pig, chicken, cow, and horse have been shown to discriminate against vitamin D₂ (99). Irradiation of 7-dehydrocholesterol or ergosterol is carried out by dissolving the steroid in an appropriate solvent, eg, peroxide-free diethyl ether. Solvents such as methanol, cyclohexane, and dioxane have also been used. The cooled solution is pumped through uv-transparent quartz reactors which permit the light from high pressure mercury lamps to impinge upon the solution (Fig. 7). The solution is recycled until the desired degree of irradiation has been achieved. This results in a mixture of unreacted 7-dehydrosterol, previtamin D, vitamin D, and irradiation by-products.

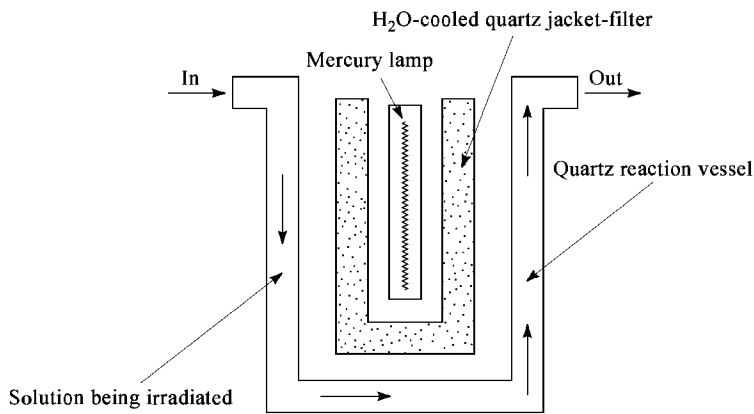


Fig. 7. Ultraviolet-light-transparent quartz reactor for sterol irradiation.

Among the light sources used for irradiation are carbon arcs, metal-corded carbon rod, magnesium arcs, and mercury-vapor lamps; the high pressure mercury lamp is most widely used. Higher yields and more favorable isomer distribution can be achieved if the frequency of light is kept at 275–300 nm. Several arrangements of the components of the simple reactor shown are possible and have been used. An important feature is a cooling jacket which controls the high temperature (ca 800°C) of the mercury-vapor lamps. Water solutions for cooling may contain salts for screening frequencies of light. Light below 275 nm can be filtered by aromatic compounds, as well as by a 5-wt % lead acetate or other inorganic salt solution. Glass filters can also be used as screens for frequencies which are outside the chemical filter ranges. Photosensitizers, eg, eosin, erythrosin, dibromodinitrofluorescein, and others, have been suggested to limit light frequencies and improve the isomer distribution.

The vitamin D resin is stabilized against oxidation by the addition of <1 wt% butylated hydroxyanisole or butylated hydroxytoluene. The solvent is then evaporated, and the unconverted sterol is recovered by precipitation from an appropriate solvent, eg, alcohol. The recovered sterol is reused in subsequent irradiations. The solvent is then evaporated to yield vitamin D resin. The resin is a pale yellow-to-amber oil that flows freely when hot and becomes a brittle glass when cold; the activity of commercial resin is 20–30 × 10⁶ IU/g. The resin is formulated without further purification for use in animal feeds. Vitamin D can be crystallized to give the USP product from a mixture of hydrocarbon solvent and aliphatic nitrile, eg, benzene and acetonitrile, or from methyl formate (100,101). Chemical complexation has also been used for purification.

In 1938, it was estimated that 7.5 × 10¹³ quanta of light were required to convert ergosterol to 1 USP unit of vitamin D₂ (102). The value was later determined to be 9.3 × 10¹³ quanta.

A flow diagram for the D₃ manufacturing process is shown in Figure 8. First, ether solution containing 7-dehydrocholesterol is recirculated through a quartz uv reactor, and the ether is distilled off. Methanol is added to the 7-dehydrocholesterol–vitamin D₃ mixture, and the remaining ether is azeotroped. The resulting solution is transferred to a crystallizer, and the 7-dehydrocholesterol is crystallized and recovered by filtration. The methanol is distilled, and the vitamin D₃ resin is heated to isomerize the pre-vitamin D to the cis-vitamin D isomer. Vitamin D₃ resin is then packaged for shipment.

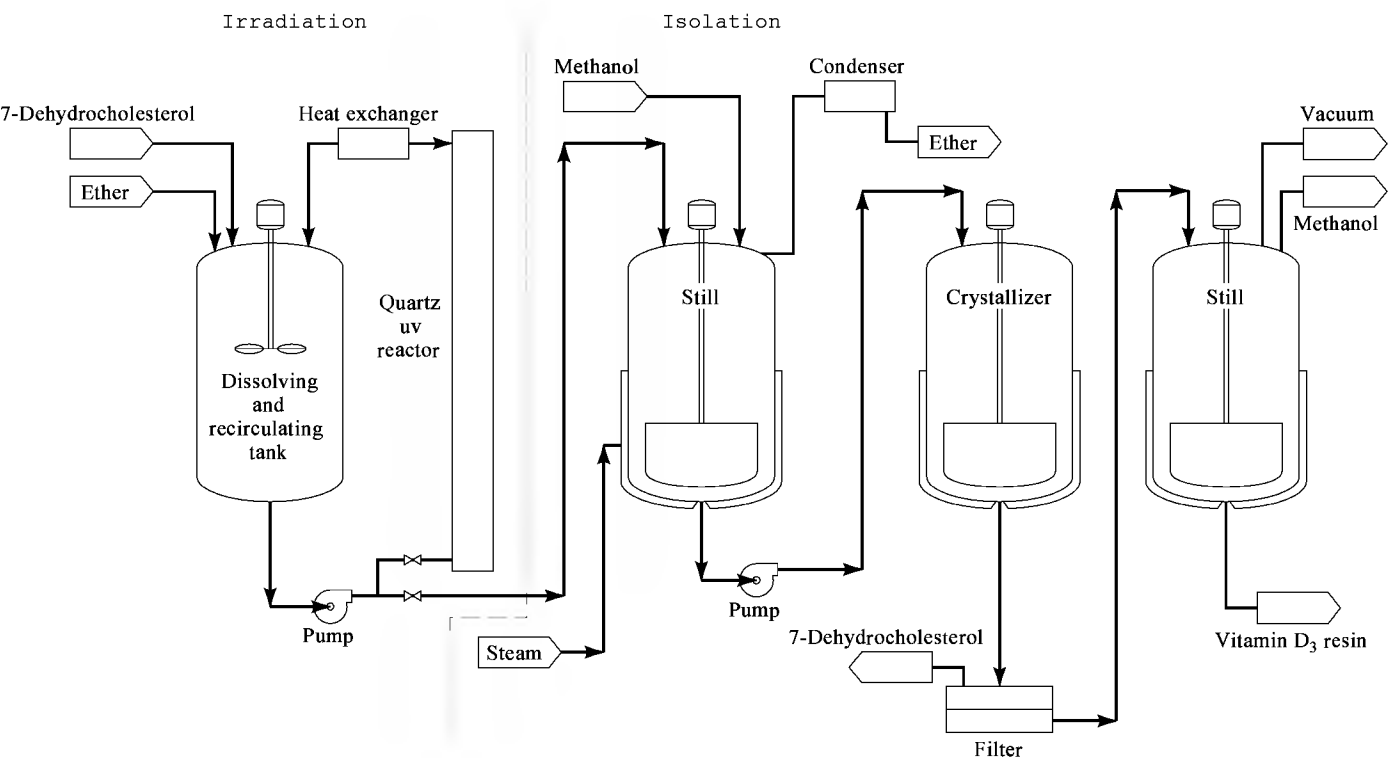
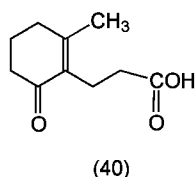


Fig. 8. Vitamin D irradiation process.

Total Synthesis. Poor yields encountered during the manufacture of vitamin D stimulated early attempts to synthesize vitamin D. In 1959 Inhoffen synthesized vitamin D₃ from 3-methyl-2-(2-carboxyethyl)-2-cyclohexenone (40), using the Wittig reaction extensively (103).



5,6-*trans*-Vitamin D₃ was prepared, followed by photochemical isomerization into vitamin D₃. Direct preparation of the previtamin and vitamins D is described in References 104 and 105.

The discovery that vitamin D₃ was metabolized to biologically active derivatives led to a significant effort to prepare 25-hydroxy vitamin D₃ and, subsequently, the 1 α -hydroxy and 1,25 dihydroxy derivatives. Initial attempts centered around modification of steroidal precursors, which were then converted to the D derivatives by conventional means.

Chemical syntheses of 25-hydroxy-vitamin D₃ were achieved by several groups of researchers (106–112). Many of these syntheses depended on the availability of the precursor 25-oxo-27-norcholesterol. Grignard reaction followed by introduction of the 7-dehydro function and irradiation allows 25-hydroxy-vitamin D₃ to be isolated (109). Fucosterol (stigmasta-5,24(28)-dien-3 β -ol) as well as bile acids, pregnenolone, and desmosterol have also been used as starting materials for the synthesis of 25-hydroxy intermediates. 24,25-Dihydroxy-vitamin D₃ [40013-87-4] was isolated and chemically characterized in 1972 (113). The first synthesis of this important D₃ catabolite was of a racemic mixture, and it was followed by the stereospecific syntheses of the two epimers (114–123). The biosynthesized material migrates exclusively with the synthetic 24(R) epimer.

The yield of the first chemical synthesis of 1,25-dihydroxy-vitamin D₃ was <0.005% (124). A key intermediate compound was 1,25-dihydroxy-cholesterol (109,125–130).

Subsequent synthesis of Vitamin D metabolites involved oxidative degradation of the vitamin D molecule to obtain the C- and D-ring portion with the intact side chain. Recombination of this molecule with an appropriate structure containing the A-ring was then carried out by a Wittig-type condensation.

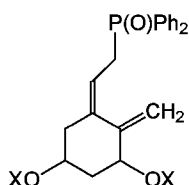
The chemical syntheses of 1,24(R),25-trihydroxy-vitamin D₃ [56142-94-0] and 1,24(S),25-trihydroxy-vitamin D [56142-95-1] were reported (131,132) in 1975. The chemical synthesis of 25,26-dihydroxy-vitamin D₃ [29261-12-9] has also been described, and it has been determined that the biologically occurring epimer is 25(R),26-dihydroxy-vitamin D₃ (117,133–135). The 23,25-dihydroxy-24-oxo metabolite has been isolated (136) as well. 1 α -Hydroxycalcitric acid (1-hydroxy-24-nor-9,10-secochola-5,7-10(19)-trien-23-oic acid) [71204-89-2], 25-hydroxy-26,23-lactone vitamin D₃ [71203-34-6], and homocalcitric acid (9,10-secochola-5,7,10(19)-trien-24-oic acid) have also been reported, and the biological activity of these metabolites has been studied (137). The synthetic chemistry associated with these and other metabolites of vitamin D₂ and vitamin D₃ is described in References 6, 40, and 138, and more recently in References 16, 51, 55, 139–141.

The discovery that vitamin D metabolites play a much larger biochemical role than just maintaining calcium homeostasis has stimulated a number of groups around the world to develop more economical chemical syntheses for the vitamin D metabolites and analogues, which might be useful in studying and treating D₃-related diseases and conditions. Many of these methods are reviewed in References 139 and 140.

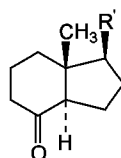
The most useful synthetic routes include the following pathways to the vitamin D structures and their derivatives:

Photochemical ring opening of 7-dehydrocholesterol derivatives which have ring A or the side chain modified (142,143).

A phosphine oxide of type (41) can be coupled with Grundman's ketone (42) to produce the D₃ skeleton (105,144–151).



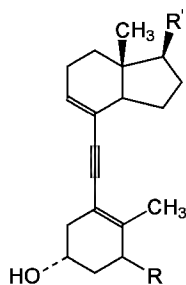
(41, X = a protective group)



(42, R' = cholesterol side chain or derivative)

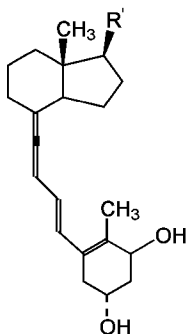
Synthetic dienynes like (43) are semihydrogenated to form previtamin D and are then rearranged to the D structure (152–155).

Vinyl allenes (44) are rearranged with heat or metal catalysis and photosensitized isomerization to produce the vitamin D triene (156–160).



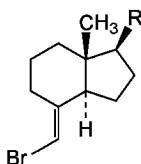
(43, R = H, OH)

R' = cholesterol side chain or derivative

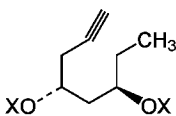


(44, R' = cholesterol side chain or derivative)

Intermolecular cross-coupling of (45) with a synthetic alkyne structure like (46) leads to the D₃ skeleton (161,162).



(45, R = cholesterol side chain or derivative)



(46, X = protective group)

Direct modification of vitamin D₂ or D₃ or its metabolites to give derivatives by a variety of synthetic methods has also been used extensively (163–171).

Whereas these preparations are extremely useful for obtaining sufficient quantities of the D₃ metabolites and analogues for study and as possible therapeutic treatments, their cost is high compared to the cost of manufacture of vitamin D₃. For this reason, the metabolites are unlikely to be successful in replacing vitamin D₃ as an ingredient in animal or human nutrition.

Biochemistry

Biochemistry. Vitamin D is introduced into the bloodstream either from the skin after natural synthesis by the irradiation of 7-dehydrocholesterol stored in the epidermis (172) or by ingestion and absorption of vitamin D₂ or vitamin D₃ through the gut wall (40). Between 60 and 80% of the vitamin introduced in the blood is taken up by the liver, where cholecalciferol is transferred from chylomicrons to a vitamin D-binding protein (DBP), an α -globulin specific for vitamin D and its metabolites but one which does not bind with previtamin D in the skin (173). Cholecalciferol is hydroxylated in the liver at the C-25 position (51,141,174). This hydroxylation occurs in the endoplasmic reticulum and requires NADPH, a flavoprotein, cytochrome P-450, Mg²⁺, and O₂ (175). 25-Hydroxylation also occurs in intestinal homogenates of chicks (176), but does not appear to occur outside the liver in mammals (177). 25-Hydroxy-cholecalciferol is the main circulating form of vitamin D₃, and its normal concentration level is 15–47 ng mL (178).

25-Hydroxy vitamin D pools in the blood and is transported on DBP to the kidney, where further hydroxylation takes place at C-1 or C-24 in response to calcium levels. 1-Hydroxylation occurs primarily in the kidney mitochondria and is catalyzed by a mixed-function monooxygenase with a specific cytochrome P-450 (52,179,180). 1 α - and 24-Hydroxylation of 25-hydroxycholecalciferol has also been shown to take place in the placenta of pregnant mammals and in bone cells, as well as in the epidermis. Low phosphate levels also stimulate 1,25-dihydroxycholecalciferol production, which in turn stimulates intestinal calcium as well as phosphorus absorption. It also mobilizes these minerals from bone and decreases their kidney excretion. Together with PTH, calcitriol also stimulates renal reabsorption of the calcium and phosphorus by the proximal tubules (51,141,181–183).

Vitamin D receptors have been identified in intestine, bone, and kidney. Receptors have also been shown, to a lesser degree, to be present in the pituitary, brain, skin, reproductive organs, and cells of the immune system, suggesting vitamin D may play a broader role in controlling cellular proliferation and differentiation (184). The nuclear receptor (genomic) for 1,25-dihydroxy vitamin D₃ has been shown to be a member of the super-family of transactivating regulators of gene transcription similar to the receptors of other steroid hormones (185). Nongenomic activation by 1,25-dihydroxy vitamin D₃ on the cellular and subcellular level has been observed and is believed to be responsible for calcium transport in tissue known as transcalcitachia (186,187).

Further side-chain oxidation of vitamin D₃ metabolites may be necessary for phosphate transport (188,189). 24,25-Dihydroxycholecalciferol is produced by kidney mitochondria, but can also be produced elsewhere (190,191). 1 α ,24,25-Trihydroxy-vitamin D₃ stimulates intestinal calcium mobilization to the extent of ca 60% of the 1 α ,25 activity and has 10% of the 1 α ,25 activity in causing bone resorption (192). It is less active in chicks and

is excreted.

More than 20 other, naturally occurring metabolites of vitamin D have been isolated and characterized, and many derivatives have been synthesized. Their function is the subject of continuing research (16,51,141,162).

Although it is being found that vitamin D metabolites play a role in many different biological functions, metabolism primarily occurs to maintain the calcium homeostasis of the body. When calcium serum levels fall below the normal range, 1 α ,25-dihydroxy-vitamin D₃ is made; when calcium levels are at or above this level, 24,25-dihydroxycholecalciferol is made, and 1 α -hydroxylase activity is discontinued. The calcium homeostasis mechanism involves a hypocalcemic stimulus, which induces the secretion of parathyroid hormone. This causes phosphate diuresis in the kidney, which stimulates the 1 α -hydroxylase activity and causes the hydroxylation of 25-hydroxy-vitamin D to 1 α ,25-dihydroxycholecalciferol. Parathyroid hormone and 1,25-dihydroxycholecalciferol act at the bone site cooperatively to stimulate calcium mobilization from the bone (see HORMONES). Calcium blood levels are also influenced by the effects of the metabolite on intestinal absorption and renal resorption.

Interaction of vitamin D and its metabolites with sex hormones has been demonstrated, particularly in birds in which the egg-laying functions combine calcium needs and reproductive activity. The metabolites of vitamin D behave as hormones. As such, they play an active role in the endocrine system, along with other hormones, to maintain the various body functions. Several biological influences of metabolites of vitamin D have been studied, including effects related to cancer (193–197), skin diseases (198–201), immunomodulatory effects (202,203), and Alzheimer's disease (204–206) (Fig. 9).

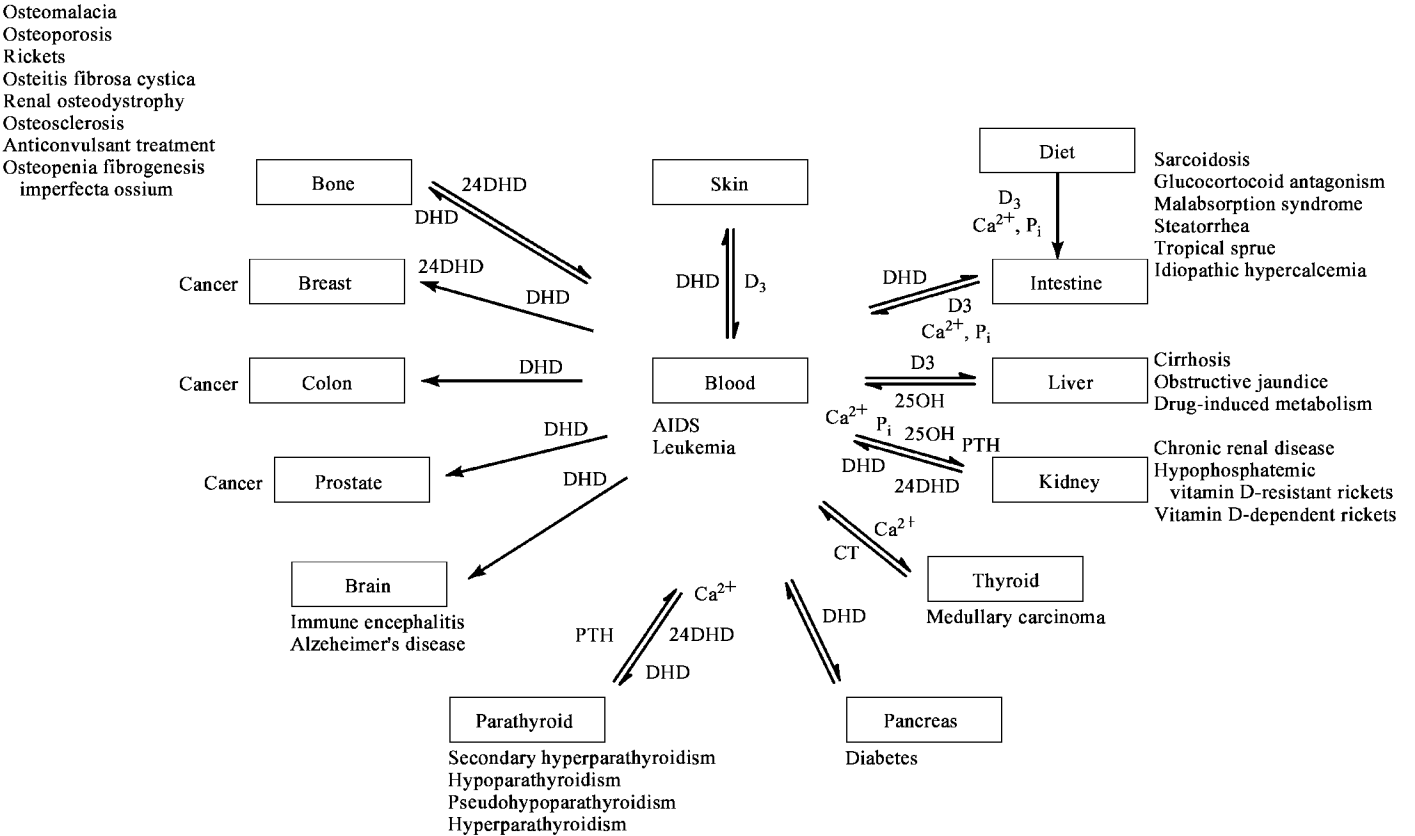


Fig. 9. Human disease states in which vitamin D has been implicated. D₃ = vitamin D₃; DHD = 1 α ,25-dihydroxyvitamin D₃; 24DHD = 24(R),25-dihydroxyvitamin D₃; 25OH = 25-hydroxyvitamin D₃; PTH = parathyroid hormone; CT = calcitonin; P_i = inorganic phosphorus.

The metabolism of vitamin D₂ follows a pathway similar to that described for Vitamin D₃.

Vitamin D₃ Deficiency. Vitamin D₃ deficiency is uncommon in normal adults. However, when it does occur, it can be serious, particularly in pregnant women. Some vitamin D₃ deficiency can occur because of a large reduction of fat intake, which decreases D₃ absorption. Strict vegetarians also risk reduced vitamin D₃ intake. Premature infants and elderly people who are exposed to minimal sunlight and consume little vitamin D₃ also have a reduced capacity to metabolize D₃ and can develop vitamin D₃ deficiency.

Clinical stresses which interfere with vitamin D₃ metabolism, can result in calcium deficiency leading to osteomalacia and osteoporosis (secondary vitamin D deficiency). These stresses include intestinal malabsorption (lack of bile salts); stomach bypass surgery; obstructive jaundice; alcoholism; liver or kidney failure decreasing hydroxylation of vitamin D₃ to active forms; inborn error of metabolism; and use of anticonverdiants that may lead to increased D₃ requirement.

People who experience the exclusion of sunlight (living in northern climates; cultures where apparel limits sunlight), as well as infants and children that are confined to bed or limited outdoor activity because of weather or illness, also can show a tendency towards vitamin D₃ deficiency.

Vitamin D deficiency in animals may be caused by the fact that the vitamin is not available to the livestock. Modern animal husbandry subjects animals to total confinement with little or no exposure to sunlight. This mandates that they be given vitamin D-fortified diets. The vitamin is sensitive to oxidation, heat, light, and minerals, and significant losses may occur in the fortified feed unless the product is adequately protected. Mycotoxins in feeds also interfere with utilization of vitamin D in feeds (207–209).

Symptoms of D₃ deficiency in animals include poor appetite, stunted growth, and weight loss; increased incidence of irritability and convulsions (tetany); some growth abnormalities; decreased egg production in poultry with reduced hatchability and thin eggshell quality; and birth of weak, dead, or deformed offspring in other animals.

For a more detailed description of symptoms, see References 4, 5, 210, and 211.

Dietary Requirements. Vitamin D is essential for growth and maintenance of good health. According to the National Research Council (NRC), the vitamin D requirement for optimum health is ca 400 IU d in humans, regardless of age. This amount of vitamin D gives ample protection from rickets, provided a sufficient amount of the other essential nutrients, including calcium and phosphorus, is supplied (212). Recommended NRC amounts of vitamin D per kilogram of feed for various species are as follows: starting and growing chicks, 200 IU; laying and breeding hens, 500 IU; turkeys, 1100 IU; ducks, 200 IU; quail, 480–900 IU; geese, 200 IU; and swine 125–220 IU. Calves require 600 IU per 100 kg of body weight (213–215). Practical feeding levels are somewhat higher, in order to assure delivery of the vitamin to the animals.

The present (ca 1997) maximum safe level of vitamin D₃ for long-term feeding in most species is four to ten times the NRC dietary requirements. Short-term (<60 d), most species can tolerate 100 times their apparent dietary requirements (210).

Animals exposed to sunlight for extended lengths of time do not require substantial dietary vitamin D. Current livestock management practices place an emphasis on high productivity, and most feed manufacturers recommend vitamin D supplementation of diets. Recommendations for practical levels of vitamin D in feeds for various animals, as recommended by feed manufacturers, are listed in Table 10.

Table 10. Practical Feeding Levels of Vitamin D, 10⁶ IU/t

Animal	Amount	Animal	Amount
<i>Poultry</i>		<i>Dairy cattle</i>	
chickens		calf starter	1–2
broilers	2–4	calf milk replacer	2–4
replacement birds	1–3	replacement heifers	1–2
layers	1–3	dry cows	1–2
breeding hens	2–4	lactating cows	2–4
turkeys		bulls	2–4
starting	3–5		
growing	2–4	<i>Beef cattle</i>	
breeding	3–5	calf starter	1–2
ducks		replacement heifers	1–2
market	1–3	feedlot	2–3
breeding	2–4	dry pregnant cows	1–2
		lactating cows	1–2
		bulls	1–2
<i>Swine</i>			
prestarter (to 10 kg)	2–3		
starter (10–35 kg)	1–2	<i>Sheep</i>	
growing-finishing (35 kg to market)	1–2	fattening lambs	2–3
gestation	1–2	breeding	1–2
lactation	2–3		
boars	2–3	<i>Other</i>	
		dogs	5–1
<i>Fish</i>		cats	1–2
trout	1–2	horses	1–2
catfish	1–2		

Historically, tickets prevention or cure was used to evaluate adequate vitamin D₃ nutrient levels. More recently, in the absence of uv light, Edwards (216) found different vitamin D₃ levels were required for the optimization of the various effects of vitamin D₃ in poultry, ie, 275 IU kg for growth, 503 IU kg for bone ash, 552 IU kg for blood plasma calcium, and 904 IU kg for rickets prevention.

Toxicity. Vitamin D toxicity was known as early as the year 1429 (217). Accidental toxicity has been reported in monkeys, dogs, horses, pigs, chinchillas, and humans, and particularly in cattle when extremely high doses of vitamin D have been used to treat milk fever.

Vitamin D intoxication causes 25-hydroxy vitamin D₃ blood levels to go from a normal of 30–50 ng mL to 200–400 ng mL. At this high level, the metabolite can compete with 1 α -25-dihydroxy vitamin D₃ for receptors in the intestine and bone and induce effects usually attributed to the dihydroxy vitamin D₃. Thus, 25-hydroxy vitamin D₃ is believed to be the critical factor in vitamin D intoxication. Vitamin D₂ is metabolized slower than vitamin D₃ and thus appears to be less toxic (218).

The overall effect in most animals is to stimulate intestinal absorption of calcium with a concomitant increase in serum calcium and a reduction in parathyroid hormone (PTH). Modest hypercalcemia allows the glomerular filtration rate to remain stable and hypercalciuria to occur because of increased filtered load of calcium and reduction of tubular resorption of calcium with reduced PTH. However, with further increases in serum calcium, the glomerular filtration rate decreases, resulting in an even more rapid increase in serum calcium and the subsequent fall in urinary calcium.

Polyuria, along with vomiting can cause extracellular fluid to be reduced, contributing to further renal function disruption. Renal dysfunction then, becomes the major contributor to loss of calcium homeostasis, and the resulting hypercalcemia during vitamin D intoxication. Loss of appetite, gastrointestinal disturbances, head and joint pain, and muscle weakness are typical clinical symptoms. Death from hypervitaminosis D is usually caused by renal failure. Cardiovascular and kidney mineralization results. Respiratory tract mineralization has also been observed, as well as calcification in the salivary gland. The severity of effects of toxic levels of vitamin D depend on the dose, functional state of the kidneys, composition of the diet, as well as differences in toxicology of the various species (210).

Vitamin D withdrawal is an obvious treatment for D toxicity (219). However, because of the 5–7 d half-life of plasma vitamin D and 20–30 d half-life of 25-hydroxy vitamin D, it may not be immediately successful. A prompt reduction in dietary calcium is also indicated to reduce hypercalcemia. Sodium phytate can aid in reducing intestinal calcium transport. Calcitonin glucagon and glucocorticoid therapy have also been reported to reduce serum calcium resulting from D intoxication (210).

Vitamin D₃ is nonirritating to rabbit eye and dermal tests. It has an oral LD₅₀ of 35–42 mg kg in rats; 42 mg kg in mice (136 mg kg IP), and 4–80 mg kg in dogs (10 mg kg IP and 5 mg kg IM and IV). It is nontoxic on inhalation.

The metabolites of vitamin D are usually more toxic than the vitamin because the feedback mechanisms that regulate vitamin D concentrations are circumvented. 25-Hydroxycholecalciferol has a one-hundredfold increase in toxicity over vitamin D₃ when fed to chicks (220) and 1 α ,25-dihydroxy vitamin D₃ is several times more toxic than the 25-hydroxy analogue. Vitamin D₂ seems to have less toxicity than vitamin D₃, a circumstance which is believed to be caused by the more efficient elimination of 25-hydroxy and the 1 α ,25-dihydroxy vitamin D₂ from the animals. Estimated safe upper dietary levels are given in Table 11.

Table 11. Estimated Safe Upper Dietary Levels^a

Species	NRC recommended dietary daily requirement, IU kg	Exposure, multiple of daily requirement	
		Long-term (<60 d)	Short-term (>60 d)
chicken	200	200	14
quail	1200	100	4
turkey	900	100	3.9

cow	300	83	3.9
catfish	1000		20
trout	1800		555
horse	400		5.5
sheep	275	91	8
pig	220	150	10

^a Ref. 210.

Disease States. Rickets is the most common disease associated with vitamin D deficiency. Many other disease states have been shown to be related to vitamin D. These can involve a lack of the vitamin, deficient synthesis of the metabolites from the vitamin, deficient control mechanisms, or defective organ receptors. The control of calcium and phosphorus is essential in the maintenance of normal cellular biochemistry, eg, muscle contraction, nerve conduction, and enzyme function. The vitamin D metabolites also have a function in cell proliferation. They interact with other factors and receptors to regulate gene transcription.

In the treatment of diseases where the metabolites are not being delivered to the system, synthetic metabolites or active analogues have been successfully administered. Vitamin D₃ metabolites have been successfully used for treatment of milk fever in cattle, turkey leg weakness, plaque psoriasis, and osteoporosis and renal osteodystrophy in humans. Many of these clinical studies are outlined in References 6, 16, 40, 51, and 141. The vitamin D receptor complex is a member of the gene superfamily of transcriptional activators, and 1,25 dihydroxy vitamin D is thus supportive of selective cell differentiation. In addition to mineral homeostasis mediated in the intestine, kidney, and bone, the metabolite acts on the immune system, β-cells of the pancreas (insulin secretion), cerebellum, and hypothalamus.

Vitamin D metabolites may therefore play an active role in diseases related to these functions, ie, leukemia, cancer (breast, colon, prostate), and autoimmune diseases (AIDS, immune encephalitis, and diabetes) (51, 141, 193–197, 202, 203).

Economic Aspects

Vitamin D is available in a variety of forms. Cod liver oil and percomorph liver oil historically were good sources of vitamin D. Recent cost increases of these materials have caused a decline in their market position. Cod liver oil sold for \$0.40–0.45 L in 1970 and as high as \$1.45 L in 1979 and \$3.43 L in 1996. The prices of the cod liver oils and of vitamin D₂ and vitamin D₃ from 1955 to 1995 are shown in Table 12.

Table 12. Price History of Vitamin D Materials^a

Year	Source	Vitamin D	
	Cod-liver oil, \$ L	Ergocalciferol (crystalline vitamin D ₂), \$ g	Cholecalciferol D ₃ , \$ g ^b
1955	1.51	0.60	0.045
1960	1.40	0.54	0.045
1965	1.71	0.48	0.045
1970	1.70	0.48	0.045
1975	5.50	0.63	0.045
1980	8.00	0.63	0.045
1985	7.25	^c	24 ^d
1990	13.00	^c	24 ^d
1995	13.00	^c	24 ^d

^a Adapted from Ref. 221.

^b 850,000 IU g unless otherwise indicated.

^c D₂ costs were not quoted after 1983.

^d 40 × 10⁶ IU g.

Vitamin D₃ is produced as a resin (20–30 × 10⁶ IU g) which sold in 1980 for ca \$1.02 10 IU. The 1979 prices of vitamin D were \$24.25 10⁰ IU (\$9.70 kg of product at 400,000 IU g) (222). In the United States in 1979, \$3.6 × 10⁶ worth of vitamin D was sold. The 1996 price for vitamin D₃ products was \$13.50–14.00 g for resin and \$24.00 g (223) for crystalline material. Vitamin D₃ feed-grade formulated products sell for approximately \$0.03–0.04 10 units of activity.

Estimates of world demand in 1979 were as high as 1300 × 10¹² IU of vitamin D₃. This was divided into thirds for Europe, the United States, and the rest of the world, respectively. Of this demand, 90% was estimated for animal-feed fortification and 10% for food and pharmaceutical uses. It is estimated that the demand will be 1500–1600 × 10¹² IU in 1997 for animal usage and 100 × 10¹² IU for human use. The United States will require approximately 500 TU (1 trillion units = 25 kg *cis*-vitamin D₃ or 17 t of resin) for animal use and 30 TU (approximately 1 t of crystalline *cis*-vitamin D₃) for human use. This represents approximately 50 t of vitamin D₃ resin yr for animal use worldwide and about 2.5 t of crystalline vitamin D₃ for human use. A substantial proportion of the vitamin D₃ is imported, and with all uses included, it is estimated that 80–90% of the sales are of vitamin D₃.

The vitamin D₂ volume is estimated at only 1200 kg yr, exclusive of Chinese sources.

Uses

Most of the vitamin D sold is synthetic. Vitamin D₂ as a concentrate or in microcrystalline forms is used in many pharmaceutical preparations, although vitamin D₃ is preferred by many manufacturers and consumers. Vitamin D₂ in the form of irradiated yeast has been used as a feed supplement for cattle, swine, and dogs, but its use has declined in favor of Vitamin D₃. In the United States, swine usage accounts for 13% and poultry for 41% of vitamin D consumption in animal agriculture. The beef and dairy industries account for 44% of vitamin D consumption, of which 22% is by dairy calves. The remaining 1% of total consumption of D vitamins is largely in prepared pet foods. European usage in 1994 was estimated at: 58% cattle, 19% swine, 20% poultry, and 3% other. Crystalline vitamin D₃ is used for medicinal preparations and formulations in the pharmaceutical industry, as well as for the fortification of fresh and evaporated milk and nonfat dry milks. Preparations based on the use of the resin are less expensive than those using crystalline vitamin D. Essentially all milk produced in the United States is fortified with vitamin D₃. Cereals and margarine are also fortified with vitamin D₃. The vitamin must be diluted to a proper dosage form, and many supplement preparations are marketed, eg, tonics, drops, capsules, and tablets; oil-based injectables are also available. Combination formulations are used widely, particularly with vitamin A for humans and with vitamins A and E for animals. Preemulsified products are used for increased bioavailability. Water-miscible formulations have also been developed (224). Solutions of vitamin D in oil or

in oil-on-dry carriers, eg, corn or flour, are used in animal feeds. These diets contain high levels of minerals, and the vitamin D formulation is not stable unless it is protected. Several stable forms are patented and sold commercially; they include beadlets or powders of dry suspensions in gelatin, carbohydrates, wax, and cellulose derivatives (89,225–228).

Animal uses employ resin in various stabilized forms at levels of 200,000; 400,000; 500,000; and 1×10^6 units of D₃ per gram. Combination products containing A and D are also available, with 650,000 units of vitamin A and 325,000 units vitamin D per gram of product being the most common dosage form.

Approximately 200 kg yr of Vitamin D₃ formulations are also marketed as rat poisons. The metabolites of vitamin D₃ and synthetic derivatives are being used or developed for treatment of osteoporosis, skin psoriasis, and other diseases in humans. 1 α -Hydroxy vitamin D₃ is being used for milk fever in cows, and 25-hydroxy vitamin D₃ has been proposed for eggshell thickness in poultry and is being marketed as an animal dietary nutritional supplement.

BIBLIOGRAPHY

"Vitamins (Vitamin D)" in *ECT* 1st ed., Vol. 14, pp. 828–849, by H. R. Rosenberg, E. I. du Pont de Nemours & Co., Inc.; in *ECT* 2nd ed., Vol. 21, pp. 549–573, by S. B. Greenbaum, Diamond Shamrock Chemical Co.; "Vitamin D" under "Vitamins" in *ECT* 3rd ed., Vol. 24, pp. 186–213, by A. L. Hirsch, A. L. Laboratories, Inc.

1. P. Karlson (chairman) and co-workers, IUPAC-IUB Joint Commission on Biochemical Nomenclature. "Nomenclature of Vitamin D recommendations 1981," *Eur. J. Biochem.* **124**, 223 (1982).
2. T. C. Angus and co-workers, *Proc. R. Soc. London, Ser. B* **105**, 340 (1931).
3. W. F. Loomis, *Sci. Am.* **223**(6), 77 (1970).
4. H. M. Frost, *Bone Dynamics in Osteoporosis and Osteomalacia*, Surgery Monograph Series, Charles C. Thomas, Publisher, Springfield, Ill., 1966.
5. H. F. DeLuca in R. B. Alfin-Slater and D. Kratchevsky, eds., *Nutrition and the Adult: Micronutrients*, Plenum Press, New York, 1980, p. 207.
6. A. W. Norman, *Vitamin D—The Calcium Homeostatic Steroid Hormone*, Academic Press, Inc., New York, 1979, p. 3.
7. E. Mellanby, *J. Physiol. (London)* **52**, L 111 (1919); *Lancet* **196**, 407 (1919).
8. K. Huldshinski, *Dtsch. Med. Wochenschr.* **45**, 712 (1919).
9. F. Reindel and H. J. Kipphan, *Justus Liebigs Ann. Chem.* **493**, 181 (1932).
10. E. Fernholz and P. N. Chakravorty, *Ber. Deutsch. Chem. Ger. A.* **67**, 2021 (1934).
11. A. Windaus and H. Grundmann, *Justus Liebigs Ann. Chem.* **524**, 295 (1936).
12. Ref. 6, pp. 39–48.
13. A. Windaus and F. Bock, *Z. Physiol. Chem. Hoppe-Seglers* **245**, 168 (1937).
14. O. Rygh, *Nature (London)* **136**, 396 (1935).
15. H. Brockmann and co-workers, *Z. Physiol. Chem. Hoppe-Seglers* **241**, 104 (1936); **245**, 96 (1937).
16. A. W. Norman, K. Schaefer, H.-G. Grigoleit, D. v. Herath, eds., *Vitamin D. Chemical, Biochemical and Clinical Update, Proceedings of the Sixth Workshop on Vitamin D*, Merano, Italy, March, 1985, Walter de Gruyter, Berlin, 1985, p. 55.
17. R. H. Wasserman, J. D. Henion, M. R. Haussler, and T. A. McCain, *Science* **194**, 853 (1976).
18. M. Peterlik and co-workers, *Biochem. Biophys. Res. Commun.* **70**, 797 (1976).
19. J. P. Simonit, K. M. L. Morris, and J. C. Collins, *J. Endocrinol.* **68**, 18 (1976).
20. H. Zucker and W. A. Rambeck, *Zentralbl. Veterinaarmed.* **28**, 436 (1981).
21. W. A. Rambeck and co-workers, *Z. Pflanzenphysiol.* **104**, 9 (1981).
22. R. H. Wasserman and co-workers, *Nutr. Rev.* **33**, 1 (1975); *J. Nutr.* **106**, 457 (1976); *Biochem. Biophys. Res. Commun.* **62**, 85 (1975).
23. Ref. 6, p. 49.
24. M. F. Hollick in Ref. 16, p. 219.
25. A. W. Norman and co-workers, *J. Biol. Chem.*, 13811–13819 (1994).
26. M. C. Dormanen and co-workers, *Biochem. Biophys. Res. Comm.* **201**, 394–401 (1994).
27. I. Nemere and co-workers, *J. Biol. Chem.* **269**, 23750–23756 (1994).
28. **U.S. Pat. 2,395,115** (Feb. 19, 1946), K. J. Goering (to Anheuser-Busch, Inc.).
29. **U.S. Pat. 2,874,171** (Feb. 17, 1959), H. A. Nelson (to Upjohn Co.).
30. **U.S. Pat. 3,006,932** (Oct. 31, 1961), J. Green, S. A. Price, and E. E. Edwin (to Vitamins, Ltd.).
31. **U.S. Pat. 2,794,035** (May 28, 1957), O. Hummel (to Zellstaff-Fabrik Waldhof).
32. **U.S. Pat. 2,865,934** (Dec. 23, 1958), R. A. Fisher (to Biofermentation Corp.).
33. **Ger. Pat. 1,252,674** (Oct. 26, 1967), K. Petzoldt,, K. Klieslich, and H. J. Koch (to Schering A.G.).
34. K. Ziegler, *Ann.* **551**, 80 (1942).
35. S. Bernstein, L. J. Benovi, L. Dorfman, K. S. Sax, and Y. Subbarow, *J. Org. Chem.* **14**, 433 (1949); **U.S. Pat. 2,498,390** (Feb. 21, 1950), S. Bernstein and K. J. Sax (to American Cyanamid Co.).
36. **U.S. Pat. 2,446,091** (May 4, 1968), J. Van Der Vliet and W. Stevens (to Hartford National Bank and Trust Co.).
37. H. Schaltagger, *Helv. Chim. Acta* **33**, 2101 (1950).
38. D. N. A. Holwerda, Doctoral Dissertation, Imperial University Leydon, 1970.
39. F. Hunziker and F. X. Mullner, *Helm. Chim. Acta* **41**, 70 (1958).
40. A. W. Norman, K. Schaefer, D. V. Herrath, and H. G. Grigoleit, eds., *Vitamin D: Chemical, Biochemical, and Clinical Endocrinology of Calcium Metabolism, Proceedings of the Fifth Workshop on Vitamin D*, Williamsburg, Va., Feb. 1982, Walter de Gruyter, Berlin, 1982, p. 1133.
41. W. G. Salmond, M. A. Serta, A. M. Cain, and M. C. Sobala, *Tetrahedron Lett.* **20**, 1683 (1977).
42. **U.S. Pat. 2,098,984** (Nov. 16, 1937), A. Windaus and F. Schenck (to Winthrop Chem. Co., Inc.).
43. L. E. Fieser, *J. Am. Chem. Soc.* **75**, 4394 (1953).
44. **U.S. Pat. 2,505,646** (Apr. 25, 1950), W. E. Meuly (to E. I. du Pont de Nemours & Co., Inc.).
45. A. Windaus, F. VonWerder, A. Luttringhaus, and E. Fernholz, *Ann.* **499**, 188 (1932).
46. L. Velluz, G. Amiard, and B. Goffinet, *Bull. Soc. Chim. France* **22**, 1341 (1955).
47. E. Havinga, R. J. DeKoch, and M. Rappoldt, *Tetrahedron* **11**, 276 (1960).
48. H. J. C. Jacobs, *Pure & Appl. Chem.* **67**(1), 63 (1995).
49. J. G. Bell and A. A. Christie, *Analyst* **99**, 385 (1974).
50. P. B. Braun, J. Hornstra, C. Knobles, E. W. M. Rutten, and C. Romer, *Acta Crystallogr.* **B29**, 463 (1973).
51. E. Havinga, *Experimentia* **29**, 1181 (1973).
52. A. W. Norman, R. Bouillon, M. Thomasset, eds., *Vitamin D. A Pluripotent Steroid Hormone: Structural Studies, Molecular Endocrinology and Clinical Applications. Proceedings of the Ninth Workshop on Vitamin D*, Orlando, Florida, May 1994, Walter de Gruyter, Berlin, 1994, p. 89.
53. J. I. Pedersen, J. G. Ghazarian, N. R. Orme-Johnson, and H. F. DeLuca, *J. Biol. Chem.* **251**, 3933 (1976).
54. A. W. Norman and co-eds., *Vitamin D; Basic Research and its Clinical Application, Proceedings of the Fourth Workshop on Vitamin D*, Berlin, Feb. 1979, Walter de Gruyter, Berlin, 1979, p. 173.
55. K. H. Hanewald, M. P. Rappoldt, and J. R. Roborgh, *Rec. Trav. Chim. Pays-Bas* **80**, 1003 (1961).

56. A. W. Norman, K. Schefer, H.-G. Grigoleit, D. Herrath, eds., *Vitamin D. Molecular, Cellular and Clinical Endocrinology. Proceedings of the Seventh Workshop on Vitamin D*, April, 1988, Walter de Gruyter, Berlin, 1988, p. 83.

57. A. Vedloop, A. L. Koevoet, and E. Havinga, *Rec. Trav. Chim. Pays-Bas* **76**, 689 (1957).

58. A. L. Hoevoet, A. Vedloop, and E. Havinga, *Rec. Trav. Chim.* **74**, 788 (1955).

59. W. G. Dauben and co-workers, *J. Am. Chem. Soc.* **80**, 4117 (1958).

60. W. H. Okamura, M. L. Hammond, A. J. C. Jacobs, and J. Von Thiegil, *Tetrahedron Lett.* **52**, 4807 (1976).

61. E. Havinga, *Chimia* **30**, 27 (1976); D. H. R. Barton and co-workers, *Chem. Comm.*, 65 (1976).

62. F. Boosman, H. J. C. Jacobs, E. Havinga, and A. Van den Gen, *Tetrahedron Lett.* **7**, 427 (1975).

63. A. W. Norman and co-eds., *Vitamin D Biochemical, Chemical and Clinical Aspects Related to Calcium Metabolism; Proceedings of the Third Workshop on Vitamin D*, Asilomar, Pacific Grove, Calif., Jan. 1977, Walter de Gruyter, Berlin, 1977, p. 15.

64. S. Kobayshi and M. Yasumura, *J. Nutr. Sci. Vitaminol.* **19**, 123 (1973); S. Kobayshi and co-workers, *J. Nutr. Sci. Vitaminol.* **26**, 545 (1980).

65. D. H. R. Barton, R. H. Heese, M. M. Pecket, and E. Rizzardo, *J. Am. Chem. Soc.* **95**, 2748 (1973).

66. K. Pfoertner and J. D. Weber, *Helv. Chim. Acta* **55**, 921, 937 (1972).

67. A. Vedloop, A. L. Koevoet, R. Van Moorselase, and E. Havinga, *Rec. Trav. Chim. Pays-Bas* **78**, 1004 (1959).

68. A. Vedloop, A. L. Koevoet, and E. Havinga, *Rec. Trav. Chim. Pays-Bas* **74**, 1125 (1955).

69. T. Kobagashi, *J. Vitaminol. (Jpn)*. **13**, 255 (1967).

70. H. H. Inhoffen, G. Quinkert, J. H. Hess, and H. M. Erdman, *Chem. Ber.* **89**, 2273 (1956).

71. C. H. Nields, W. C. Russell, and A. Zimmerli, *J. Biol. Chem.* **136**, 73 (1940).

72. J. B. Wilkie, S. W. Jones, and O. L. Kline, *J. Amer. Pharm. Assoc. Sci. Ed.* **47**, 395 (1958).

73. S. C. Eyley and D. H. Williams, *J. Chem. Soc. Chem. Comm.*, 858 (1975).

74. E. C. Snoeren, M. R. Daha, J. Lugtenburg, and E. Havinga, *Rec. Trav. Chim. Pays-Bas* **89**, 261 (1970).

75. W. G. Dauben and R. B. Phillips, *J. Am. Chem. Soc.* **104**, 355 (1982).

76. V. Malatesta, C. Willis, and P. A. Hackett, *J. Am. Chem. Soc.* **103**, 6781 (1981).

77. **U.S. Pat. 3,661,939** (May 9, 1972), M. Toyoda, Y. Tawara (to Nisshin Flour Milling Co., Ltd.).

78. P. S. Chen and H. B. Bosman, *J. Nutr.* **83**, 133 (1964).

79. Code of Federal Regulations, Title 21-Food and Drugs, Part 582.5923, p. 541, April 1, 1995.

80. *Chron. WHO* **3**, 747 (1965).

81. *The United States Pharmacopeia 23 (USP 23-NF 18)*, The United States Pharmacopeial Convention, Inc., Rockville, Md., 1995.

82. P. Cuniff, ed., *Official Methods of Analysis of the International Association of Official Analytical Chemists*, 16th ed., Washington, D.C., 1995.

83. F. Mulder, *J. Assoc. Off. Anal. Chem.* **60**, 989 (1977).

84. H. Hofsass, N. J. Alicino, A. L. Hirsch, L. Amieka, and L. D. Smith, *J. Assoc. Off. Anal. Chem.* **61**, 774 (1978).

85. F. Mulder and co-workers, *J. Assoc. Off. Anal. Chem.* **64**, 61 (1981); **65**, 1228 (1982); **68**, 822 (1985).

86. Ref. 6, p. 64.

87. J. Carol, *J. Pharm. Sci.* **50**, 451 (1961).

88. W. W. Morris, Jr., J. B. Wilkie, S. W. Jones, and L. Friedman, *Anal. Chem.* **34**, 381 (1962).

89. R. Strohecker and H. M. Henning, *Vitamin Assay-Tested Methods*, CRC Press, Cleveland, Ohio, 1966, p. 281.

90. U.S. Pats. 2,777,797 and 2,777,798 (Jan. 15, 1977), M. Hochberg and M. L. MacMillan (to Nopco Chemical Co.).

91. J. R. Evans, *Clin. Chim. Acta* **42**, 167 (1972).

92. T. K. Murray, K. C. Day, and E. Kodicek, *Biochem. J.* **98**, 29P (1966).

93. L. V. Aviolo and S. W. Lee, *Anal. Biochem.* **16**, 193 (1966).

94. D. Sklan, P. Budowski, and M. Katz, *Anal. Biochem.* **56**, 606 (1973).

95. D. O. Edlund and F. A. Filippini, *J. Assoc. Anal. Chem.* **56**, 1374 (1973).

96. T. G. Hogness, A. E. Sidwell, Jr., and F. P. Zscheile, Jr., *J. Biol. Chem.* **120**, 239 (1937).

97. W. Huber, G. W. Ewing, and J. Kriger, *J. Am. Chem. Soc.* **67**, 609 (1945).

98. R. H. Wasserman and co-workers, *Vitam. Horm. (N.Y.)* **32**, 299 (1974).

99. R. L. Horst and co-workers, Ref. 55, p. 93.

100. **U.S. Pat. 3,334,118** (Aug. 1, 1967), K. Schaff, S. Schmuhler, and H. C. Klein (to Nopco Chemical Co.).

101. **U.S. Pat. 3,665,020** (May 23, 1972), R. Marbet (to Hoffmann-LaRoche, Inc.).

102. R. S. Harris, J. W. M. Bumker, and L. M. Moser, *J. Am. Chem. Soc.* **60**, 2579 (1938).

103. H. H. Inhoffen and co-workers, *Chem. Ber.* **91**, 2309 (1958); H. H. Inhoffen and co-workers, *Chem. Ber.* **92**, 1564 (1959).

104. T. M. Dawson, J. Dixon, P. S. Littlewood, B. Lythgoe, and A. K. Saksena, *J. Chem. Soc. C*, 2960 (1971).

105. B. Lythgoe, M. E. N. Nambudiry, and J. Tideswell, *Tetrahedron Lett.* **31**, 3685 (1977).

106. J. A. Campbell, D. M. Squires, and J. C. Babcock, *Steroids* **13**, 567 (1969).

107. S. J. Halkes and N. P. VanVliet, *Rec. Trav. Chim. Pays-Bas* **88**, (1969).

108. J. W. Blunt and H. F. DeLuca, *Biochemistry* **8**, 671 (1969).

109. M. Morisaki, J. Rubio-Lightbourn, and N. Ikekawa, *Chem. Pharm. Bull.* **21**, 457 (1973).

110. J. J. Partridge, S. Faber, and M. R. Uskokovic, *Helv. Chim. Acta* **57**, 764 (1974).

111. **U.S. Pat. 4,172,076** (Apr. 4, 1977), A. Hirsch and J. Pikel (to Diamond Shamrock, subsequently assigned to A. L. Labs, Jan. 1, 1982).

112. **U.S. Pat. 4,226,770** (Oct. 7, 1980), E. Kaiser.

113. M. H. Holick and co-workers, *Biochemistry* **11**, 4251 (1972).

114. Y. Tanaka, H. Frank, H. F. DeLuca, N. Koizumi, and N. Ikekawa, *Biochemistry* **14**, 3293 (1975).

115. Y. Tanaka, H. F. DeLuca, N. Ikekawa, M. Morisaki, and N. Koizumi, *Arch. Biochem. Biophys.* **170**, 620 (1975).

116. H.-Y. Lam, H. K. Schnoes, H. F. DeLuca, and T. C. Chen, *Biochemistry* **12**, 4851 (1973).

117. M. Seki, J. Rubio-Lightbourn, J. Morisaki, and N. Ikekawa, *Chem. Pharm. Bull.* **21**, 2783 (1973).

118. J. Redel, P. Bell, F. Delbarre, and E. Kodicek, *C. R. Acad. Sci.* **278**, 529 (1974).

119. J. Redel and co-workers, *J. Steroid Biochem.* **6**, 117 (1975).

120. N. Ikekawa, M. Morisaki, N. Koizumi, Y. Kato, and T. Takeshita, *Chem. Pharm. Bull.* **23**, 695 (1975).

121. G. Milhaud, M.-L. Labat, J. Redel, *C. R. Acad. Sci.* **279**, 827 (1974).

122. M. Seki, N. Koizumi, M. Morisaki, and N. Ikekawa, *Tetrahedron Lett.*, 15 (1975).

123. J. J. Partridge, V. Toome, and M. R. Uskokovic, *J. Am. Chem. Soc.* **98**, 3739 (1976).

124. E. J. Semmler, M. F. Holick, H. K. Schnoes, and H. F. DeLuca, *Tetrahedron Lett.*, 4147 (1972).

125. K. Ochi and co-workers, *J. Chem. Soc. Perkin I*, 165 (1979).

126. T. A. Narwid, J. F. Blunt, J. A. Jacobelli, and M. R. Uskokovic, *Helv. Chim. Acta* **57**, 781 (1974).

127. J. Rubio-Lightbourn, M. Morisaki, and N. Ikekawa, *Chem. Pharm. Bull.* **21**, 1854 (1973).

128. M. Morisaki, K. Bannai, and N. Ikekawa, *Chem. Pharm. Bull.* **21**, 1853 (1973).

129. M. Morisaki, J. Rubio-Lightbourn, N. Ikekawa, and T. Takeshita, *Chem. Pharm. Bull.* **21**, 2568 (1973).

130. H. E. Paaren, D. E. Hamer, H. K. Schnoes, and H. F. DeLuca, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2080 (1978).

131. N. Ikekawa and co-workers, *Biochem. Biophys. Res. Commun.* **62**, 485 (1975).

132. J. J. Partridge, S. J. Shivey, E. G. Baggiolini, B. Hennessy, and M. Uskokovic, in A. W. Norman and co-eds., *Vitamin D: Biochemical, Chemical and Clinical Aspects Related to Calcium Metabolism*, Walter de Gruyter, Berlin, 1977, p. 47.
133. H.-Y. Lam, H. K. Schnoes, and H. F. DeLuca, *Steroids* **25**, 247 (1975).
134. J. Redel, P. Bell, F. Delbarre, and E. Kodicek, *C. R. Acad. Sci.* **276**, 2907 (1973).
135. J. Redel and co-workers, *Steroids* **24**, 463 (1974); J. Redel, N. Bazely, Y. Tanaka, and H. F. DeLuca, *FEBS Lett.* **94**, 228 (1978).
136. A. W. Norman and co-workers, *Biochemistry* **22**, 1798 (1983).
137. H. F. DeLuca and H. K. Schnoes, in A. W. Norman and co-eds., *Vitamin D: Recent Basic Advances and Their Clinical Application*, Walter de Gruyter, Elmsford, N.Y., 1979.
138. D. E. M. Lawson, *Vitamin D*, Academic Press, Inc., New York, 1978.
139. H. Dai and G. H. Posner, *Synthesis*, 1383 (1994).
140. W. H. Okamura and C. D. Zhu, *Chemical Reviews* **95**(c), 1877 (1995).
141. A. W. Norman, R. Bouillon, and M. Thomasset, eds., *Vitamin D, Structure, Function Analysis and Clinical Application, Proceedings of the Eighth Workshop on Vitamin D*, Paris, July 5, 1991, Walter de Gruyter, Berlin, 1991.
142. D. H. R. Barton, R. H. Hesse, M. M. Pechet, and E. Rizzardo, *J. Chem. Soc., Chem. Commun.*, 203–204 (1974).
143. D. H. R. Barton, R. H. Hesse, M. Pechet, and E. Rizzardo, *J. Am. Chem. Soc.* **95**, 2748 (1973).
144. B. Lythgoe and co-workers, *Tetrahedron Lett.*, 3863–3866 (1975).
145. B. Lythgoe and co-workers, *J. Chem. Soc., Perkin Trans.*, **1** 590–595 (1978).
146. B. Lythgoe, T. A. Moran, M. E. N. Nambudiry, and S. Ruston, *J. Chem. Soc., Perkin Trans.*, **1**, 2386–2390 (1976).
147. B. Lythgoe and co-workers, *J. Chem. Soc., Perkin Trans.*, **1**, 387–395 (1978).
148. J. V. Frosch, I. T. Harrison, B. Lythgoe, A. K. Saksena, *J. Chem. Soc., Perkin Trans.*, **1**, 2005–2009 (1974).
149. E. G. Baggiolini and co-workers, *J. Org. Chem.* **51**, 3098–3108 (1986).
150. J. DeSchrijver and R. J. De Clercq, *Tetrahedron Lett.* **34**, 4369 (1993).
151. G. H. Posner and C. M. Kinter, *J. Org. Chem.* **55**, 3967–3969 (1990).
152. L. Castedo, J. L. Mascarenas, A. Mourino, and L. A. Sarandeses, *Tetrahedron Lett.* **29**, 1203–1206 (1988).
153. L. Castedo, A. Mourino, and L. A. Sarandeses, *Tetrahedron Lett.* **29**, 1203–1206 (1988).
154. S. A. Barrack, R. A. Gibbs, and W. J. Okamura, *J. Org. Chem.* **53**, 1790–1796 (1988).
155. J. M. Aurrecoechea and W. H. Okamura, *Tetrahedron Lett.* **28**, 4947–4950 (1987).
156. W. H. Okamura, J. M. Aurrecoechea, R. A. Gibbs, and A. W. Norman, *J. Org. Chem.* **54**, 4072–4083 (1989).
157. E. M. VanAlstyne, A. W. Norman, and W. H. Okamura, *J. Am. Chem. Soc.* **116**, 6207–6216 (1994).
158. M. L. Hammond and co-workers, *J. Am. Chem. Soc.* **100**, 4907 (1978).
159. P. Condran and co-workers, *J. Am. Chem. Soc.* **102**, 6259 (1980).
160. W. H. Okamura and co-workers, in Ref. 53, pp. 5–12.
161. B. M. Trost, J. Dumas, and M. Villa, *J. Am. Chem. Soc.* **114**, 9836–9845 (1992).
162. K. Nagasawa, Y. Zako, H. Ishihara, and I. Shimizu, *Tetrahedron Lett.* **32**, 4937–4940 (1991).
163. M. Sheves and Y. Mazur, *J. Am. Chem. Soc.* **97**, 6249–6250 (1975).
164. H. E. Paaren, H. F. DeLuca, and H. K. Schnoes, *J. Org. Chem.* **45**, 3253–3258 (1980).
165. M. Kabat and co-workers, *Tetrahedron Lett.* **32**, 2343–2346 (1991).
166. S. R. Wilson, A. M. Venkatesan, C. E. Augelli-Szafran, and A. Yasmin, *Tetrahedron Lett.* **32**, 2339–2342 (1991).
167. D. R. Andrews and co-workers, *J. Org. Chem.* **51**, 1635–1637 (1986).
168. H. H. Inhoffen, G. Quinkert, and S. Schutz, *Chem. Ber.* **90**, 1283–1286 (1957).
169. H. H. Inhoffen, G. Quinkert, H.-J. Hess, and H. Hirschfeld, *Chem. Ber.* **90**, 2544–2553 (1957).
170. D. R. Andrews, D. H. R. Barton, R. H. Hesse, and M. M. Pechet, *J. Org. Chem.*, **51**, 4819–4828 (1986).
171. L. Vanmaele, P. J. De Clercq, and M. Vandewalle, *Tetrahedron Lett.* **41**, 141–144 (1985).
172. M. F. Holick, *Am. J. Clin. Nutr.* **60**, 619–630 (1994).
173. M. F. Holick and co-workers, *Endocrinology* **135**, 655 (1994).
174. G. Ponchon, A. L. Kennan, and H. F. DeLuca, *J. Clin. Invest.* **48**, 1273 (1969).
175. T. C. Madhok and H. F. DeLuca, *Biochem. J.* **184**, 491 (1979); E. Axen, T. Bergman, and K. Wikvall, *J. Steroid Biochem. Mol. Biol.* **51**, 97 (1994).
176. S. A. Holick, M. F. Holick, T. E. Tavel, H. K. Schnoes, and H. F. DeLuca, *J. Biol. Chem.* **251**, 397 (1976).
177. E. B. Olson, Jr., J. C. Knutson, M. H. Bhattacharyya, and H. F. DeLuca, *J. Clin. Invest.* **57**, 1213 (1976).
178. J. A. Eisman, R. M. Shepard, and H. F. DeLuca, *Anal. Biochem.* **80**, 198 (1977).
179. D. R. Fraser and E. Kodicek, *Nature (London)* **228**, 764 (1970).
180. J. G. Ghazarian, C. R. Jefcoate, J. C. Knutson, W. H. Orme-Johnson, and H. F. DeLuca, *J. Biol. Chem.* **249**, 3026 (1974).
181. L. T. Boyle, L. Miravet, R. W. Gray, M. F. Holick, and H. F. DeLuca, *Endocrinology* **90**, 605 (1972).
182. T. C. Chen, L. Castillo, M. Korycka-Dahl, and H. F. DeLuca, *J. Nutr.* **104**, 1056 (1974).
183. H. F. DeLuca and co-workers, in C. Hansch, P. G. Sammes, and J. B. Taylor, eds., *Comprehensive Medicinal Chemistry*, Vol. 3, Pergamon Press, Oxford, U.K., 1991, p. 1129.
184. A. W. Norman and co-workers, *J. Steroid Biochem. Mol. Biol.* **41**, 231 (1992).
185. K. E. Lowe, A. C. Maiyor, and B. W. Norman, *Crit. Rev. Eukar. Gene. Exp.* **2**, 65–109 (1992).
186. D. T. Baran, *J. Cell. Biochem.* **270**, 303 (1994).
187. D. W. Beno and co-workers, *J. Biol. Chem.* **270**, 3542 (1995).
188. R. Kumar, D. Hamden, and H. F. DeLuca, *Biochemistry* **15**, 2420 (1976).
189. D. Hamden, R. Kumar, M. F. Holick, and H. F. DeLuca, *Science* **193**, 493 (1976).
190. M. F. Holick and co-workers, *Biochemistry* **11**, 4251 (1972).
191. Y. Tanaka, L. Castillo, H. F. DeLuca, and N. Ikekawa, *J. Biol. Chem.* **252**, 1421 (1977).
192. M. F. Holick and co-workers, *J. Biol. Chem.* **248**, 6691 (1973).
193. J. N. M. Heersche and J. A. Kanis, eds., *Bone and Mineral Research*, Vol. 8, Elsevier Science B.V., Amsterdam, the Netherlands, 1994, p. 45.
194. M. Gross and co-workers, *J. Bone Miner. Res.* **1**, 457 (1986).
195. T. V. Wijngaarden and co-workers, *Cancer Res.* **54**, 5711 (1994).
196. R. J. Skowronski, D. M. Pechl, and D. Feldman, *Endocrinology* **136**, 20 (1995).
197. M. Inaba and co-workers, *Blood* **82**, 53 (1993).
198. J. A. MacLaughlin and co-workers, *Proc. Natl. Acad. Sci.* **82**, 5409 (1985).
199. S. Morimoto and co-workers, *Calif. Tissue Int.* **39**, 209 (1986).
200. S. Morimoto and co-workers, *Calif. Tissue Int.* **38**, 119 (1986).
201. K. J. Kragballe, *J. Cell. Biochem.* **49**, 46 (1992).
202. J. M. Lemire, *J. Cell. Biochem.* **49**, 26 (1992).
203. E. P. Amento, *Steroids* **49**, 55 (1987).
204. M. S. Saporito and co-workers, *Brain Res.* **633**, 189 (1994).

205. M. S. Saporito and co-workers, *Exp. Neurol.* **123**, 295 (1993).

206. S. Carswell, *Exp. Neurol.* **124**, 36 (1993).

207. H. Kohler and co-workers, *Zentralbl. Veterinaarmed. Reihe B.* **25**, 89 (1978).

208. F. H. Bird, *Poult. Sci.* **57**, 1293 (1978).

209. B. Jedek and coworkers, *Zentralbl. Veterinaarmed. Reihe B.* **25**, 29 (1978).

210. *Vitamin Tolerances of Animals*, National Research Council, National Academy Press, Washington, D.C., 1987, p. 10.

211. Y. H. Hull, *Encyclopedia of Food Science and Technology*, Vol. 4, John Wiley & Sons, Inc., New York, 1991, Pts. 1–8, p. 2713.

212. *Recommended Dietary Allowances*, 7th ed., National Academy of Science, National Research Council, Washington, D.C., 1968, p. 1964.

213. *Nutritient Requirements of Poultry*, 9th ed., National Academy of Science, National Research Council, Washington, D.C., 1994, p. 15.

214. *Nutritient Requirements of Swine*, 9th ed., National Academy of Science, National Research Council, Washington, D.C., 1988, p. 35.

215. *Nutritional Requirements of Dairy Cattle*, 3rd ed., National Academy of Science, National Research Council, Washington, D.C., 1966, p. 1349.

216. H. Edwards, "Factors Influencing Leg Disorders in Broilers," p. 21 in (Proceedings of the Maryland Nutrition Conference, March 23, 1995, University of Maryland, College Park, Maryland).

217. W. Putscher, *Z. Kinderheilkd.* **48**, 269 (1929).

218. D. O. Harrington and E. H. Page, *J. Am. Vet. Med. Assoc.* **182**, 1358 (1983).

219. K. Diem and C. Lentner, eds., *Scientific Tables*, Ciba Geigy, Ltd., Basel, Switzerland, 1971, p. 464.

220. R. L. Morrissey and co-workers, *J. Nutr.* **107**, 1027 (1977).

221. *Oil, Paint and Drug Reporter*, **167**, No. 2, p. 23 (1955) and subsequent issues.

222. *Chemical Economics Handbook*, Stanford Research Institute, Menlo Park, Calif., 1980.

223. *Chem. Mktg. Rep.* **251**, No. 5, p. 39 (1997).

224. **U.S. Pat. 2,417,299** (Mar. 11, 1947), L. Freedman and E. Green (to U.S. Vitamin Corp.).

225. **U.S. Pat. 2,702,262** (Feb. 15, 1955), Burley and A. E. Timrech (to Charles Pfizer and Co., Inc.).

226. **U.S. Pat. 2,827,452** (Mar. 18, 1978), H. Schlenk, D. M. Sand, and J. A. Tillotson (to University of Minnesota).

227. **U.S. Pat. 3,067,104** (Dec. 4, 1962), M. Hochberg and C. Ely (to Nopco Chemical Co.).

228. **U.S. Pat. 3,143,475** (Aug. 4, 1964), A. Koff and R. F. Widmer (to Hoffmann-LaRoche, Inc.).

General References

Refs. 5, 6, 16, 51, 62, and 141 are also general references.

C. E. Bells, in W. H. Sebrell and R. S. Harris, eds., *The Vitamins*, 1st ed., Vol. 2, Academic Press, New York, 1954, p. 132.

G. F. Combs and H. F. DeLuca, *The Vitamins: Fundamental Aspects in Nutrition and Health*, Academic Press, Inc., 1992, p. 151.

E. D. Collins and A. W. Norman, in L. J. Macklin, ed., *Handbook of Vitamins*, 2nd ed., Marcel Dekker, New York, 1991, Chapt. 2, p. 59.

H. F. DeLuca in R. B. Olfen-Slater and D. Kritchevsky, eds., *Nutrition and the Adult: Micronutrients*, Plenum Press, New York, 1986, p. 205.

W. Friedrich, *Vitamins*, Walter de Gruyter, New York, 1988, p. 141.

A. W. Norman, in M. Brown, ed., *Present Knowledge in Nutrition*, 6th ed., International Life Sciences Institute–Nutrition Foundation, Washington, D.C., 1990, Chapt. 12, p. 108.

Arnold L. Hirsch
Alpharma Inc.

VITAMIN E

Vitamin E was first described in 1922 and the name was originally applied to a material found in vegetable oils. This material was found to be essential for fertility in rats. It was not until the early 1980s that symptoms of vitamin E deficiency in humans were recognized. Early work on the natural distribution, isolation, and identification can be attributed to Evans, Burr, and Emerson (University of California) and Mattill and Olcott (University of Iowa). Subsequently a group of substances (Fig. 1), which fall into either the family of tocopherols or tocotrienols, were found to act like vitamin E (1–4). The structure of α -tocopherol was determined by degradation studies in 1938 (5).

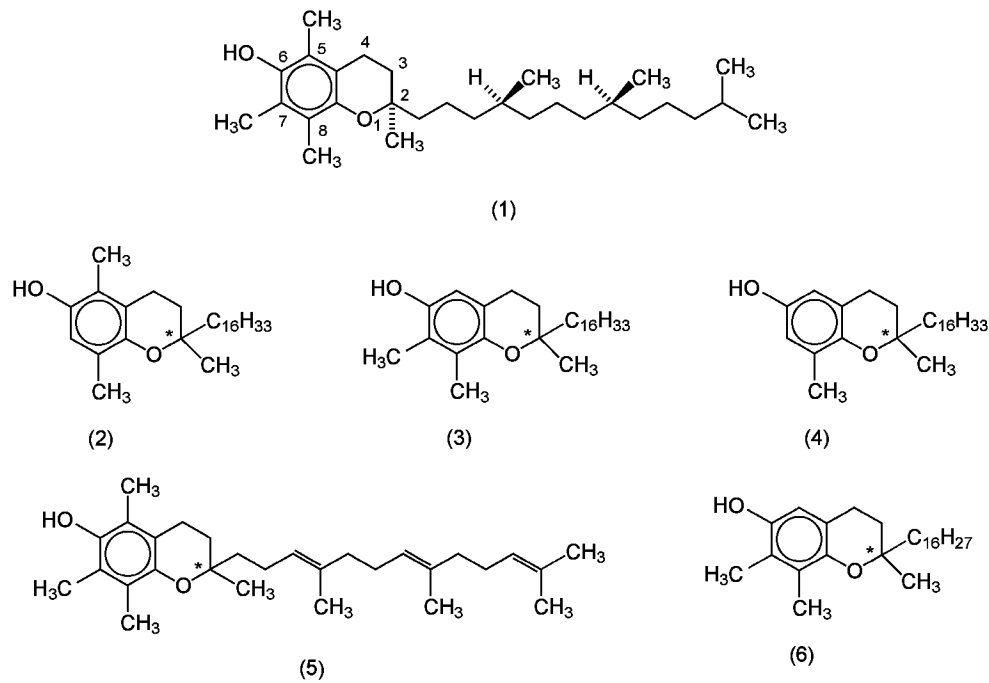


Fig. 1. The four naturally occurring tocopherols (α -tocopherol, RRR [59-02-9] *all-rac* [2074-53-5] (1); β -tocopherol [148-03-8] (2); γ -tocopherol [54-28-4] (3); δ -tocopherol [119-13-1] (4)), α -tocotrienol [1721-51-3] (5), and β -tocotrienol [14101-61-2] (6) where asterisks denote asymmetric centers and the absolute configuration of RRR- α -tocopherol is shown.

Chemical, Biological, and Physical Properties

The structures (see Fig. 1) of all vitamin E compounds are characterized by a 6-chromanol ring with a C₁ side chain. As a result of three asymmetric carbons, there are eight possible optical isomers. However, it is the configuration around the 2-position on the ring which determines, to the largest extent, biological activity (6). Natural vitamin E is characterized by all R stereochemistry. The tocopherols are distinguished by a saturated phytol side chain and the tocotrienols are characterized by an unsaturated side chain. The isomers within the two families differ by the extent and position of the methylation on the ring. Although tocopherols and tocotrienols are chemically related, they have varying degrees of vitamin E activity depending on their structure and stereochemistry (Table 1).

Table 1. Relative Activity of Vitamin E Forms^a

Substance	Vitamin E activity
<i>all-rac</i> - α -tocopheryl acetate	1.00
RRR- α -tocopheryl acetate	1.36
<i>all-rac</i> - α -tocopheryl acid succinate	0.89
RRR- α -tocopheryl acid succinate	1.21
<i>all-rac</i> - α -tocopherol	1.10
RRR- α -tocopherol	1.49
β -tocopherol	0.50
γ -tocopherol	0.35
α -tocotrienol	0.30
γ -tocotrienol	0.10

^a Refs. 7–9.

The biological activity of various vitamin E forms was established by the fetal resorption assay in rats and is assumed to be applicable to humans. The results of some human studies may indicate that the ratio of 1.36 underestimates the biological activity of the RRR form relative to the *all-rac* form of α -tocopheryl acetate (10–12). α -Tocopherol and α -tocopheryl acetate are viscous oils. The fat solubility of α -tocopherol and α -tocopheryl acetate is characteristic of the family of compounds. Both are readily soluble in ethanol, chloroform, and acetone but are insoluble in water. Other physical properties of α -tocopherols and their

acetate esters are given in Table 2.

Table 2. Physical Properties of Tocopherols

Property	<i>all-rac</i> - α -Tocopherol	RRR- α -Tocopherol	<i>all-rac</i> - α -Tocopheryl Acetate	RRR- α -Tocopheryl -Acetate
CAS Registry Number	[2074-53-5]	[59-02-9]	[7695-91-2]	[58-95-7]
color	colorless to pale yellow			
form	viscous oil			
bp, °C	200–220 (0.1 mm)		224 (0.3 mm)	
sp gr ²⁵ ₂₅	0.947–0.958	0.950–0.964	0.950–0.964	
<i>n</i> ²⁰ _D	1.5030–1.5070	1.4940–1.4985	1.4940–1.4985	
mol wt	430.69	430.69	472.73	472.73
uv absorption maxima, nm	292–294	292–294	285.5	285.5
<i>E</i> _{1 cm} ^{1%} (ethanol)	71–76	71–76	40–44	40–44

Although they are generally stable to heat and alkali, and acids in the absence of oxygen, free tocopherols and tocotrienols can be oxidized by atmospheric oxygen in the presence of light, unsaturated fatty acids, heat, or metal ions. The reported oxidation products of α -tocopherol are shown in Figure 2. These products can be readily synthesized by oxidizing α -tocopherol with such reagents as nitric acid, silver nitrate, ferric chloride, or benzoyl peroxide. As a result of the ease of oxidation, significant amounts of vitamin E (tocopherols) may be lost during food processing (qv). The severity of the processing, ie, cooking, canning, and baking, determines the amount of tocopherol remaining in foods and edible oils. Storage conditions can also impact the losses of vitamin E in foods and animal feeds (see FEEDS AND FEED ADDITIVES). The stability of α -tocopherol is significantly enhanced by esterification. Although the acetate ester does not have any inherent antioxidant activity, it is bioactive and is the commercially important form because of its greater stability.

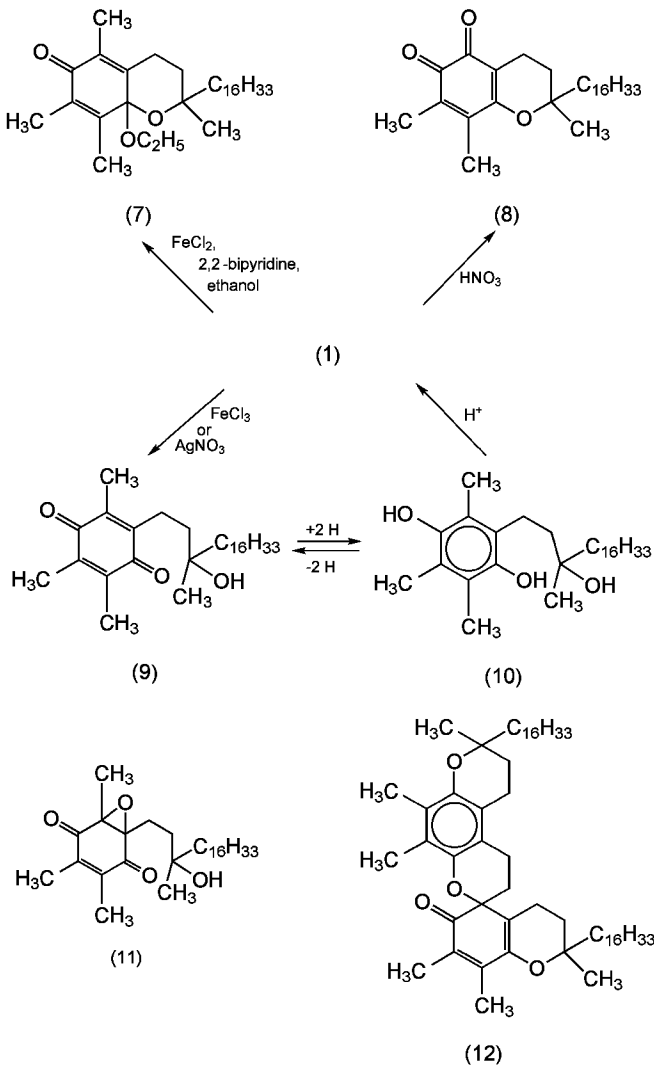


Fig. 2. Oxidation products of α -tocopherol: tocopherethoxide [511-72-8] (7), tocored [17111-16-9] (8), α -tocoquinone [7559-04-8] (9), α -tocohydroquinone [14745-36-9] (10), α -tocoquinone-2,3-oxide [35499-91-3] (11), and α -tocopherol dimer [1604-73-5] (12).

Natural Occurrences and Sources

Unesterified tocopherols are found in a variety of foods; however, concentration and isomer distribution of tocopherols vary greatly with source. Typically, meat, fish, and dairy contain <40 mg 100 g of total tocopherols. Almost all (>75%) of this is α -tocopherol for most sources in this group. The variation in the content of meat and dairy products can be related to the content of the food ingested by the animal. A strong seasonal variation can also be observed. Vegetable oils contain significant levels of γ -, β -, and δ -tocopherol, along with α -tocopherol (Table 3).

Table 3. Vitamin E Content of Common Oils, mg/100 g^a

Oil source	Total tocopherols	α-Tocopherol	γ-Tocopherol	β-Tocopherol	δ-Tocopherol
soybean	14	8–10	59		26
corn	24	10–16	60	5	2
safflower	41	40	17		1
canola	23	19	43		
cottonseed	48	39–44	39		
sunflower	49	49	5		0.8
coconut	1	0.5–1			0.6
palm	29	20–26	32		7
olive	10	5–10			
peanut		13	22		2
rapeseed		18	17		1
wheatgerm		133	26	71	27

⁰⁰³ Refs. 13 and 14.

Vegetable oils, typically soybean, are important feedstocks for the commercial production of the RRR forms of vitamin E.

Synthesis

Although all four tocopherols have been synthesized as their *all-nat* forms, the commercially significant form of tocopherol is *all-nat*-α-tocopheryl acetate. The commercial processes in use are based on the work reported by several groups in 1938 (15–17). These processes utilize a Friedel-Crafts-type condensation of 2,3,5-trimethylhydroquinone with either phytol (16), a phytyl halide (7,16,17), or phytadiene (7). The principal synthesis (Fig. 3) in current commercial use involves condensation of 2,3,5-trimethylhydroquinone (13) with synthetic isophytol (14) in an inert solvent, such as benzene or hexane, with an acid catalyst, such as zinc chloride, boron trifluoride, or orthoboric acid oxalic acid (7,8,18) to give the *all-nat*-α-tocopherol (15a). Free tocopherol is protected as its acetate ester (15b) by reaction with acetic anhydride. Purification of tocopheryl acetate is readily accomplished by high vacuum molecular distillation and rectification (<1 mm Hg) to achieve the required USP standard.

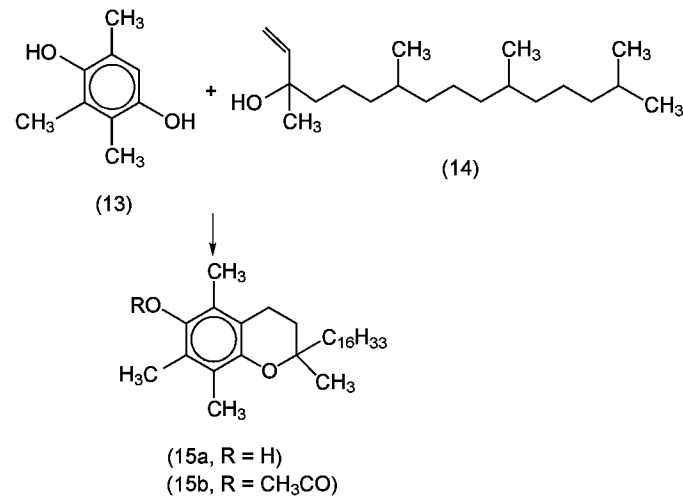


Fig. 3. Synthesis via trimethylhydroquinone [700-13-0] (TMHQ) (13) and isophytol [505-32-8] (14) of α-tocopherol (15a) and α-tocopherol acetate (15b).

Trimethylhydroquinone can be synthesized from various isomeric trimethylphenols such as 2,4,6- or 2,3,6-trimethylphenol (Fig. 4). When starting with 2,3,6-trimethylphenol (16), this can be accomplished by catalytic oxidation using oxygen and cupric halide catalysts in a solvent such as water–alcohol, 2-methoxyethanol, acetone, or *N,N*-dimethylformamide (19,20). The resulting benzoquinone (17) can then be hydrogenated catalytically using Pd or Pt in an inert solvent, such as methanol, isobutanol, or toluene (21) to yield 2,3,5-trimethylhydroquinone (13). When 2,4,6-trimethylphenol (18) is used, 4-hydroxy-2,4,6-trimethyl-2,5-cyclohexadien-1-one (19) is obtained by oxidation with sodium hypochlorite (22) or oxygen and base (23). This compound can then be rearranged by base to trimethylhydroquinone (22,23). The trimethylphenols can be obtained by methylation of phenol.

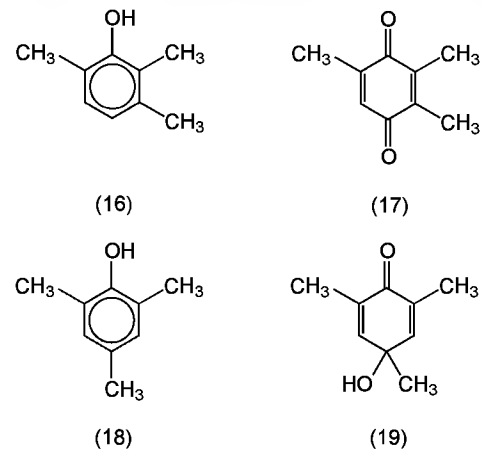


Fig. 4. Starting materials and intermediates in the TMHQ syntheses: 2,3,6-trimethylphenol [2416-94-6] (16), 2,3,5-trimethylquinone [935-92-2] (17), 2,4,6-trimethylphenol [527-60-6] (18), and 4-hydroxy-2,4,6-trimethyl-2,5-cyclohexadien-1-one [2875052-9] (19).

The isophytol side chain can be synthesized from pseudoionone (Fig. 5) using chemistry similar to that used in the vitamin A synthesis (9). Hydrogenation of pseudoionone (20) yields hexahydropseudoionone (21) which can be reacted with a metal acetylide to give the acetylenic alcohol (22). Rearrangement of the adduct of (22) with isopropenylmethyl ether yields, initially, the allenic ketone (23) which is further transformed to the C₁₈-ketone (24). After reduction of (24), the saturated ketone (25) is treated with a second mole of metal acetylide. The acetylenic alcohol (26) formed is then partially hydrogenated to give isophytol (14).

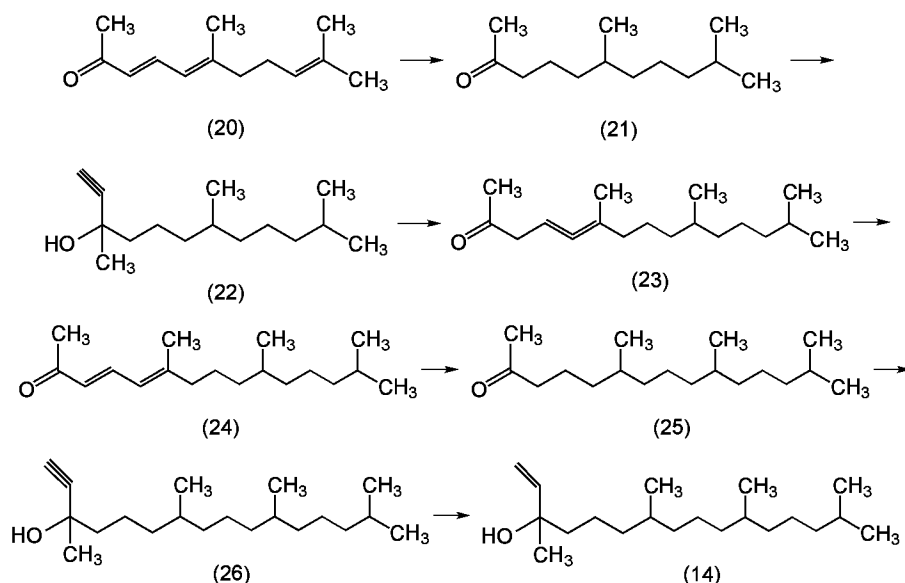


Fig. 5. Synthesis of isophytol (14): pseudoionone [141-10-6] (20), hexahydropseudoionone [1604-34-8] (21), C₁₅-acetylenic alcohol [1604-35-9] (22), C₁₈-allenic ketone [16647-10-2] (23), C₁₈-diene ketone [1604-32-6] (24), C₁₈-saturated ketone [16825-16-4] (25), and C₂₀-acetylenic alcohol [29171-23-1] (26).

Although the eight stereoisomers of α -tocopherol have been synthesized (24), these preparations are at present only of academic interest. However, RRR- α -tocopherol and the other natural forms of vitamin E can be isolated from deodorizer distillates produced as by-products of vegetable oil processing. This requires a series of steps built around key molecular distillations which are required to increase the purity of the RRR- α -tocopherol and mixed tocopherol fractions for use in commercial preparations. A typical process would involve a distillation of alkali-treated soybean oil at high vacuum (<1 mm Hg) in a continuous molecular still. This minimizes thermal degradation. The distillate, which contains α -, γ -, and δ -tocopherols, is then cooled to low temperature ($\sim 10^\circ\text{C}$) in a solvent to remove impurities such as sterols. The other impurities, such as fatty acids, can be removed by saponification. Further molecular distillation is required to produce a fraction containing high levels (>60%) of tocopherols. The sterols and fatty acids can be sold.

The yield of the more active RRR- α -tocopherol can be improved by selective methylation of the other tocopherol isomers or by hydrogenation of α -tocotrienol (25,26). Methylation can be accomplished by several processes, such as simultaneous haloalkylation and reduction with an aldehyde and a hydrogen halide in the presence of stannous chloride (27), aminoalkylation with ammonia or amines and an aldehyde such as paraformaldehyde followed by catalytic reduction (28), or via formylation with formaldehyde followed by catalytic reduction (29).

The *all-rac* forms of β -, γ -, and δ -tocopherols can be synthesized using the same condensation reaction as used for *all-rac*- α -tocopherol. To synthesize *all-rac*- β -tocopherol, 2,5-dimethylhydroquinone instead of trimethylhydroquinone is condensed with isophytol. For *all-rac*- γ - and δ -tocopherol, 2,3-dimethylhydroquinone and methylhydroquinone are used, respectively.

Although apparently not commercially important, fermentation (qv) processes have been reported for the production of tocopherols. *Aspergillus niger* (30), *Lactobacter* (31), *Englena gracilis* Z. (32,33), and *Mycobacterium* (34) have been shown to produce (RRR)- α -tocopherol. In the case of *Englena*, titers of 140–180 mg/L have been reported.

Physiological Effects and Requirements

The symptoms of vitamin E deficiency in animals are numerous and vary from species to species (13). Although the deficiency of the vitamin can affect different tissue types such as reproductive, gastrointestinal, vascular, neural, hepatic, and optic in a variety of species such as pigs, rats, mice, dogs, cats, chickens, turkeys, monkeys, and sheep, it is generally found that necrotizing myopathy is relatively common to most species. In humans, vitamin E deficiency can result from poor fat absorption in adults and children. Infants, especially those with low birth weights, typically have a vitamin E deficiency which can easily be corrected by supplements. This deficiency can lead to symptoms such as hemolytic anemia, reduction in red blood cell lifetimes, retinopathy, and neuromuscular disorders.

Vitamin E can also act as an antioxidant (qv) in animals and humans alone or in combination with vitamin C (qv). Both are good free-radical scavengers with the vitamin C acting to preserve the levels of vitamin E (35). Vitamin E in turn can preserve the levels of vitamin A in animals (13). It has been shown that vitamin E reduces the incidence of cardiovascular disease (36–39). This most likely results from the antioxidant property of the vitamin which inhibits the oxidation of low density lipoproteins (LDLs) (40–42). The formation of the oxidized LDLs is considered important in decreasing the incidence of cardiovascular disease (43).

The recommended daily allowance for vitamin E ranges from 10 international units (1 IU = 1 mg *all-rac*- α -tocopheryl acetate) for children under 4 years of age to 30 IU for adults and children over 4 years of age. This should be adequate to prevent vitamin E deficiency in humans. High levels enhance immune responses in both animals and humans. Requirements for animals vary from 3 USP units/kg diet for hamsters to 70 IU/kg diet for cats (13). The complete metabolism of vitamin E in animals or humans is not known. The primary excreted breakdown products of α -tocopherol in the body are glucuronides of tocopheronic acid (27) (Fig. 6). These are derived from the primary metabolite α -tocopheryl quinone (9) (see Fig. 2) (44,45).

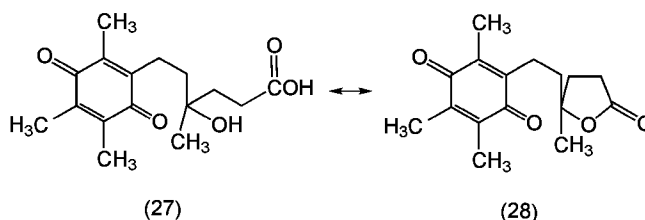


Fig. 6. α -Tocopheronic acid [1948-76-1] (27) and its lactone, α -tocopheronolactone [3121-68-4] (28).

Vitamin E is considered nontoxic at levels up to 3200 mg day. It has not been found to be teratogenic, mutagenic, or carcinogenic at doses below 1 g kg of body weight. Vitamin E can heighten the effect of vitamin K deficiency (coagulation defect) or anticoagulation therapy. A recent study, however, indicates that it may be the oxidation product α -tocoquinone that causes the effect (46).

Economic Aspects

World production in the late 1990s of both natural and synthetic forms of vitamin E is estimated at 22,000 metric tons and growth is expected to keep pace with increasing need. The 1993 U.S. production was 14,096 metric tons (47) with an additional 1080 metric tons from imports. The principal U.S. producers of the natural form are Eastman Chemical Company, Archer Daniels Midland Company, and Henkel, and of synthetic vitamin E, Hoffmann-La Roche and BASF. International producers include Hoffmann-La Roche, BASF, Eisai, and Rhône-Poulenc.

The price of synthetic vitamin E decreased steadily until about 1990 when prices began to increase. The price of the natural form, although higher than the synthetic form, has approximately paralleled the synthetic form (Table 4).

Table 4. Bulk Prices of Pharmaceutical-Grade α -Tocopheryl Acetate, \$/kg^a

Product	1948	1951	1954	1960	1970	1982	1991	1994	1995
(RRR)-form	250	250	185	122	68	68	57	59	59
<i>all rac</i> -form	750	350	136	90	50	27	23	34	34

^a Refs. 48–51.

Analytical Methods and Specifications

Specifications and standards for various vitamin E forms and preparations for use in pharmaceutical applications are given in the *United States Pharmacopeia* (52). All products should contain not less than 96.0% or more than 102.0% of the appropriate form. The products must be labeled to indicate both the chemical and stereochemical forms contained in the product.

Label claims for tocopherol levels in preparations can be based on milligrams or International Units. Only the RRR or *all-rac*- α -tocopherol and its esters can be claimed. No vitamin E activity can be claimed for the tocotrienols and the non- α -tocopherols. International Units are also used in some reference books and compendia, eg, *Food Chemicals Codex* (40,53), which is of particular importance for specifications for food fortification.

All formulations of vitamin E must show low acidity, and contain not more than 0.004% heavy metals (reported as Pb) and not more than 10 ppm Pb. Formulations that contain RRR- α -tocopherol must have a specific rotation of +24° for the oxidation product with alkaline potassium ferricyanide.

Analysis of vitamin E can be done by a variety of methods, depending on the form and level present and the preparation in which the form is found (54). For pure or highly concentrated forms, this is accomplished by reaction with Emmerie-Engel reagent (2,2'-bipyridine (α,α' -dipyridyl) and ferric chloride) to give a red color. This color results from the combination of bipyridine with ferrous ions formed from the reduction of the ferric ions by the tocopherol and is directly proportional to the amount of tocopherol present. Analysis of α -tocopherol in forms which contain low levels, such as vegetable oils, foods, or feeds, is difficult because the colorimetric method is nonspecific and significant sample preparation is involved to remove interferences such as other tocopherols, tocotrienols, and β -carotene. The AOAC Official Methods of Analysis describes a packed column gas chromatographic method for the analysis of tocopherol isomers in mixed concentrates (41). This method separates α - and δ -tocopherols as discrete peaks, but β - and γ -tocopherols elute as a combined peak.

Numerous high pressure liquid chromatographic techniques have been reported for specific sample forms: vegetable oils (55,56), animal feeds (57,58), sera (59,60), plasma (61,62), foods (63,64), and tissues (63). Some of the methods require a saponification step to remove fats, to release tocopherols from cells, and or to free tocopherols from their esters. All require an extraction step to remove the tocopherols from the sample matrix. The methods include both normal and reverse-phase hplc with either uv absorbance or fluorescence detection. Application of supercritical fluid (qv) chromatography has been reported for analysis of tocopherols in marine oils (65).

Bioassay Method. A modification of the Evans resorption–gestation method can be used to determine α -tocopherol activity in supplements. Although the method is time-consuming, when carefully performed, it can produce reasonably accurate results. The method requires raising female rats on a vitamin E-free diet and mating them with normal males. Unless a vitamin E supplement containing more than 0.3–1.0 mg (depends on methodology) of α -tocopherol is administered in the first 10–12 days of pregnancy, the embryos die and are reabsorbed without apparent harm to the mother. If the test supplement contains greater than the threshold dose of α -tocopherol, the pregnancy proceeds normally.

Applications and Product Forms

Both α -tocopherol and its esters are constituents of multivitamin and single-dose nutrient capsules or liquid dietary supplements. Supplements for human use range from a few milligrams in multivitamin preparations to 500–1000 mg in single-dose supplements. Specialty items, such as ointments, creams, salves, and suppositories containing vitamin E provide other outlets for α -tocopherol. Tocopherols have significant application in cosmetics (qv) and dermatology. Tocopherols can be used in topical cream or oil forms to treat chronic skin diseases (66), reduce scarring in wound healing (67), reduce inflammation (68), and protect against uv radiation (69). Tocopherols are also incorporated into cosmetic formulations to reduce nitrosamine and nitrosamide formation from amines and amides also present in the formulation (70).

Animal feeds consume approximately 40% of the commercial production of α -tocopherol acetate. Although tocopherol is normally found in feed, the poultry, beef and dairy cattle, lamb, and swine industries use vitamin E in supplements and concentrates to replace tocopherol lost during feed processing and storage. The acetate ester is added to the feeds as either a dry, granular, free-flowing, nondusting powder containing 44,000–276,000 units per kilogram (20,000–125,000 units per pound) or as a high potency oil concentrate (23–100% α -tocopheryl acetate).

Foods are not typically fortified with vitamin E because of the lack of signs of deficiency in the general population. Fortification does occur for infant formulas as a supplement and cereals to replace processing losses. Tocopherols are finding additional applications as antioxidants in foods as the less expensive butylated hydroxytoluene (BHT) and butylated hydroxyanisole (BHA) used are being prohibited in more and more countries. Foods such as lards and citrus oils which do not contain appreciable levels of natural tocopherols can be protected with α -tocopherol at 0.02% levels (weight based on lipid phase) (9). The meat from turkeys and chickens is less prone to rancidity when refrigerated or frozen if the animal's diet contains vitamin E. The meat from pigs fed vitamin E supplements also has better storage (71) and color (72) stability than meat from pigs with normal diets. It has also been reported that veal has better storage stability (73) and beef has better color stability (74) when vitamin E supplements are used in the animal feeds. Tocopherol can be used in bacon and similar foods to prevent the formation of nitrosamines. The traditional antioxidants, such as BHT and phosphites, used in high density polyethylene (HDPE) and polypropylene (PP), can be replaced by *all-rac*- α -tocopherol (75–77).

BIBLIOGRAPHY

"Vitamin E" in *ECT* 1st ed., Vol. 14, pp. 849–858, by P. L. Harris and N. D. Embree, Eastman Kodak Co.; "Vitamins (Vitamin E)" in *ECT* 2nd ed., Vol.

21, pp. 574–585, by D. C. Herting, Tennessee Eastman Co.; in *ECT* 3rd ed., Vol. 24, pp. 214–227, by D. C. Herting, Eastman Kodak Co.

1. J. F. Pennock, F. W. Hemming, and J. D. Kerr, *Biophys. Res. Commun.* **17**, 542 (1964).
2. International Union of Nutritional Sciences, *Nutr. Abstr. Rev.* **48**, 831 (1978).
3. R. H. Bunnell, J. Keating, A. Quaresimo, and G. K. Parman, *Am. J. Clin. Nutr.* **17**, 1 (1965).
4. J. G. Bieri and R. P. Evarts, *J. Am. Diet. Assoc.* **62**, 147 (1973).
5. E. Fernholz, *J. Am. Chem. Soc.* **60**, 1741 (1938).
6. H. Weiser and M. Vecchi, *Int. J. Vitam. Nutr. Res.* **52**, 351–370 (1982).
7. S. Kasparak, in L. J. Machlin, ed., *Vitamin E: A Comprehensive Treatise*, Marcel Dekker, Inc., New York, 1980, pp. 7–65.
8. S. Kijima, in M. Mino, H. Nakamura, A. I. Diplock, and H. J. Kayden, eds., *Vitamin E: Its Usefulness in Health and Curing Diseases*, Japan Societies Press, Tokyo, Japan, and S. Karger AG, Basel, Switzerland, 1993, p. 7.
9. D. Bhatia and co-workers, "Vitamins," in the *Encyclopedia of Food Technology*, Vol. 4, parts 1–8, John Wiley & Sons, Inc., New York, 1991.
10. G. W. Burton and K. U. Ingold, in L. Packer and G. Fuchs, eds., *Vitamin E in Health and Disease*, Marcel Dekker, Inc., New York, 1993, pp. 329–344.
11. R. V. Acuff, S. S. Thedford, N. N. Hidioglou, A. M. Papas, and T. Odom, *Am. J. Clin. Nutr.* **60**, 397–402 (1994).
12. H. J. Kayden and M. G. Traber, *T. Lipid Res.* **34**, 343–358 (1993).
13. L. J. Machlin, in L. Machlin, ed., *Handbook of Vitamins*, Marcel Dekker, Inc., New York, 1990, pp. 99–144.
14. H. T. Slover, *Lipids* **6**, 291–296 (1971).
15. P. Karrer, H. Fritzsche, B. H. Ringier, and H. Solomon, *Helv. Chim. Acta* **21**, 520 (1938).
16. F. Bergel, A. Jacob, A. R. Todd, and T. S. Work, *Nature* **142**, 36 (1938).
17. L. I. Smith, H. E. Ungnade, and W. W. Prichard, *Science (Washington, D.C.)* **88**, 37 (1938).
18. **Ger. Pat. 92-4208477** (1993), P. Grafen, H. Kiefer, and H. Jaedicke (to BASF AG).
19. **U.S. Pat. 5,041,572** (Aug. 20, 1991), U. Hearcher, B. Jessel, B. Bockstiegel, P. Grafen, and H. Laas (to BASF AG).
20. **Ger. Pat. 2221624** (Nov. 1972), W. Brenner (to Hoffmann-LaRoche AG).
21. **Eur. Pat. Appl. 87-115070** (1988), T. Yui and A. Ito (to Mitsubishi Gas Co. Inc.).
22. **Jpn. Pat. 62,263,136(86-104088)** (1987), T. Kunitomi and H. Tamai (to Kuraray Co. Ltd.).
23. Y. Ichikawa, Y. Yamanaka, N. Suzuki, T. Naruchi, O. Kobayashi, and H. Tsuruta, *Ind. Eng. Chem. Prod. Res. Dev.* **18**(4), 373–375 (1979).
24. N. Cohen, C. G. Scott, C. Neukom, R. J. Lopresti, G. Weber, and G. Saucy, *Helv. Chim. Acta* **64**, 1158 (1981).
25. J. F. Pennock, F. W. Hemming, and J. D. Kerr, *Biochem. Biophys. Res. Commun.* **17**, 542 (1964).
26. H. May, P. Schindl, R. Ruegg, and O. Isler, *Helv. Chim. Acta* **46**, 963 (1963).
27. **U.S. Pat. 2,486,539** (Nov. 1, 1949), L. Weisler (to Distillation Products).
28. **U.S. Pat. 2,519,863** (Aug. 22, 1950), L. Weisler (to Eastman Kodak Co.).
29. **U.S. Pat. 2,592,628** (Apr. 15, 1952), L. Weisler (to Eastman Kodak Co.).
30. Y. Kawai, M. Otaka, M. Kavkio, Y. Oeda, N. Inoue, and H. Shinano, *Hokkaido Daigaku Suisangakubu Kenkyo Ibo* **45**(1), 26–31 (1994).
31. H. Ariga, M. Okazaki, and M. Yamada, *Rakuna Kagaku, Shokuhin no Kenkyu* **41**(5) (1992).
32. Y. Tani and S. Osuka, *Agric. Biol. Chem.* **53**(9), 2313–2318 (1989).
33. Y. Tani and H. Tsumura, *Agric. Biol. Chem.* **53**(2), 302–315 (1989).
34. **Jpn. Pat. 57,163,490(82,163,490)** (Oct. 1982), (to Shiseido Co., Ltd.).
35. E. R. Pacht, H. Kaseki, J. R. Mohammed, D. G. Cornwell, and W. B. Davis, *J. Clin. Invest.* **77**, 789 (1986).
36. N. G. Stephens, A. Parsons, P. M. Schofield, F. Kelly, K. Cheeseman, M. J. Mitchinson, and M. J. Brown, *Lancet*, **347**, 781–786 (1996).
37. M. J. Stampfer, C. H. Hennekens, J. E. Manson, G. A. Colditz, B. Rosner, and W. C. Willett, *N. Eng. J. Med.* **328**, 1444–1449 (1993).
38. E. B. Rimm, M. J. Stampfer, A. Ascherio, E. Giovannucci, G. A. Colditz, and W. C. Willett, *N. Eng. J. Med.* **328**, 1450–1456 (1993).
39. K. G. Losonczy, T. B. Harris, and R. J. Havlik, *Am. J. Clin. Nutr.* **64**, 190–196 (1996).
40. H. M. G. Princen, G. Poppel, C. Vogelzang, R. Buytenhek, and F. J. Kok, *Arterioscler. Thromb.* **12**, 554–562 (1992).
41. H. Esterbauer, M. Dieber-Rotheneder, M. Striegel, and G. Waeg, *Am. J. Clin. Nutr.* **53**, 314s–321s (1991).
42. I. Jialal, C. J. Fuller, and B. A. Huet, *Arterioscler. Thromb. Vasc. Biol.* **15**, 190–198 (1995).
43. D. Steinberg, *Circulation*, **84**, 1420–1425 (1991).
44. C. Drevon, in Ref. 4, pp. 75–76.
45. C. K. Chow, *World Rev. Nutr. Diet.* **45**, 133–166 (1985).
46. P. Dowd and Z. B. Zheng, *Proc. Natl. Acad. Sci. USA* **92**, 8171–8175 (1995).
47. Synthetic Organic Chemicals, *United States Production and Sales, 1993*, USITC Publication 2810, U.S. Government Printing Office, Washington, D.C., 1994, pp. 3–191.
48. *Chem. Mark. Rep.* **221**(6), 47 (1982).
49. *Chem. Mark. Rep.* **240**(25), 32 (1991).
50. *Chem. Mark. Rep.* **245**(5), 31 (1994).
51. *Chem. Mark. Rep.* **247**(17), 45 (1995).
52. *The United States Pharmacopeia XXIII (USP XXIII-NF XVIII)*, The United States Pharmacopeia Convention, Inc., Rockville, Md., 1995, p. 1631.
53. Food and Nutrition Board, National Research Council, *Food Chemicals Codex*, 3rd ed., National Academy Press, Washington, D.C., 1981, p. 330.
54. K. Helrich, ed., *Official Methods of Analysis of the Association of Official Analytical Chemists*, 15th ed., Association of Official Analytical Chemists, Inc., Arlington, Va., 1990, pp. 1070–1079.
55. K. Abe, Y. Yuguchi, and G. Katsui, *J. Nutr. Sci. Vitaminol.* **21**, 183–188 (1975).
56. E. Guzman-Contreras and F. C. Strong III, *J. Agric. Food Chem.* **30**, 1109–1112 (1982).
57. H. Cohan and M. R. LaPointe, *J. Assoc. Off. Anal. Chem.* **63**(3), 1254–1257 (1980).
58. C. H. McMurray and W. J. Blanchflower, *J. Chrom.* **176**, 488–492 (1979).
59. L. Jansson, B. Nilsson, and R. Lingren, *J. Chrom.* **181**, 242–247 (1980).
60. W. J. Driskell, J. Neese, C. C. Bryant, and M. M. Bashor, *J. Chrom.* **231**, 439–444 (1982).
61. C. H. McMurray and W. J. Blanchflower, *J. Chrom.* **178**, 525–531 (1979).
62. J. Lehman and H. Martin, *Clin. Chem.* **28**(8), 1784–1787 (1982).
63. J. N. Thompson and G. Hatina, *J. Liq. Chrom.* **2**(3), 327–344 (1979).
64. U. Manz and K. Philipp, *Int. J. Vit. Nutr. Res.* **51**, 342–348 (1981).
65. A. Staby, C. Borch-Jensen, S. Balchen, and J. Mollerup, *Chromatographia* **39**(11–12), 697–705 (1994).
66. W. Nikolowski, in *Vitamins* (3) Editones <Roche>, F. Hoffmann-LaRoche & Co. Ltd., Basle, Switzerland, (1973).
67. M. Ehdich, H. Traver, and T. Hunt, *Annal. Surg.* **175**(2) (Feb. 1972), pp. 31–36.
68. M. Kamimura, *J. Vitaminol.* **18**, 201–209 (1972).
69. W. A. Pryor, *Free Radicals in Biology*, Academic Press, Inc., New York, 1976, Chapt. 1.
70. W. Mergens and E. DeRiter, *Cosmetic Technol.* (Jan. 1980).
71. F. J. Monahan, D. J. Buckley, P. A. Morrissey, P. B. Lynch, and J. I. Gray, *Food Sci. Nutr.* **42F**, 203 (1990).

72. A. Asgher, J. I. Gray, A. M. Booser, E. A. Goma, M. M. Abonzied, and E. R. Miller, *J. Sci. Food Agric.* **57**, 31 (1991).
73. R. Ellis, W. I. Kimato, J. Bitman, and L. F. Edmonson, *J. Am. Chem. Soc.* **51**, 4 (1974).
74. F. B. Shorland, J. O. Igene, A. M. Pearson, J. W. McGuffy, and A. E. Aldridge, *J. Agric. Food Chem.* **29**, 863 (1981).
75. D. Burdick, S. Laermer, S. Young, and P. Zambetti, "A New Primary Antioxidant System for Polyolefins," presented at *Additives '95*, Clearwater, Fla., Feb. 22–24, 1995.
76. S. F. Laermer and P. F. Zambetti, *J. Plast. Film Sheet.* **8**, 228–248 (1992).
77. Y. C. Ho, K. L. Yam, S. S. Young, and P. F. Zambetti, *J. Plast. Film Sheet.* **10**, 194–212 (1994).

General Reference

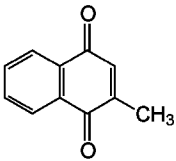
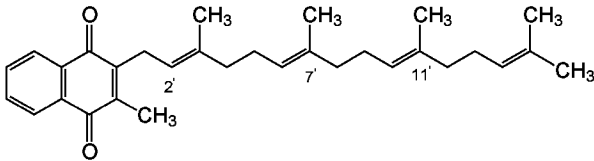
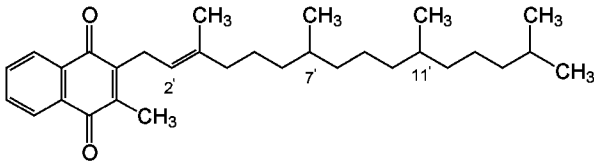
L. J. Machlin, ed., *Vitamin E: A Comprehensive Treatise*, Marcel Dekker, Inc., New York, 1980.

Robert Casani
Hoffmann-La Roche Inc.

VITAMIN K

Vitamin K represents a class of substances which contain the 2-methyl-1,4-naphthoquinone moiety and are characterized by their antihemorrhagic properties. Vitamin K was first discovered by Dam in 1929 during experiments with chicks on a lipid-deficient diet (1). Dam coined the term *Koagulations vitamin* to describe this new substance (2,3). In the late 1930s, several groups identified the causative agent responsible for the antihemorrhagic properties of alfalfa (4–7). Structural determination of vitamin K₁ (1) was the result of work from several laboratories (8,9). Stereochemical assignment came much later from Mayer (10). Additional studies isolated and characterized related materials. Vitamin K₂ (2) was first found in decaying fish meal (9); it is a crystalline compound and is distinguished by its repeating isoprenyl units. Vitamin K₃ [58-27-5] (3) is the simplest form of the vitamin. The chemical name and common names for some important forms of the vitamin follow.

name	vitamin K ₁
CAS Registry Number	[84-80-0]
chemical name	1,4-naphthalenedione, 2-methyl-3-(3,7,11,15-tetramethyl-2-hexadecenyl)-, [R-[R*,R*-(E)]]-
common name	phyloquinone, phytomenadione, phytonadione
name	vitamin K ₂₍₂₀₎
CAS Registry Number	[863-61-6]
chemical name	1,4-naphthalenedione, 2-methyl-3-(3,7,11,15-tetramethyl-2,6,10,14-hexadecatetraenyl)-, (E,E,E,E)-
common name	menaquinone 4, menaquinone K4, MK-4
name	vitamin K ₃
CAS Registry Number	[58-27-5]
chemical name	1,4-naphthalenedione, 2-methyl
common name	menadione; menaquinone-0; synkay



The K vitamins are named after the original vitamin and the number of carbon atoms in the side chain. Using this convention, vitamin K₂₍₂₀₎ is so named because it contains 20 carbon atoms in the chain. In the biological literature, vitamin K₂ is frequently referred to as menaquinone and is further designated by the number of isoprene units in the side chain. For example, vitamin K₂₍₂₀₎ is also called menaquinone-4 or MK-4. Vitamin K₃ is also referred to as menadione or MK-0.

Vitamin K is typically found in green, leafy vegetables such as cabbage, broccoli, and spinach at levels of 95–200 µg 100 g of fresh vegetables. Cauliflower at a level of 136 µg 100 g also represents an excellent source of dietary vitamin K. Additionally, animal sources such as liver and eggs provide good sources of vitamin K (11).

Chemical/Physical Properties

Vitamin K compounds are yellow solids or viscous liquids. The natural form of vitamin K₁ is a single diastereoisomer with 2'(*E*), 7'(*R*), 11'(*R*) stereochemistry. The predominant commercial form of vitamin K₁ is the racemate and a 2'(*E*) (*Z*) mixture. Table 1 lists some physical and spectral properties of vitamin K₁.

Table 1. Physical and Spectral Properties of Vitamin K₁^a

Item	Vitamin K ₁	Vitamin K _{2 30}	Vitamin K _{2 35}	Vitamin K ₃
color	yellow, viscous oil	yellow crystals	yellow crystals	bright yellow crystals
melting point, °C	20	50	54	105–107
molecular weight	450.68	580.9	649.02	172.17
molecular formula	C ₃₁ H ₄ O ₂	C ₄₁ H ₅ O ₂	C ₄ H ₄ O ₂	C ₁₁ H ₈ O ₂
spectrophotometric data, λ _{max} , nm	242, 248, 260, 269, 325	243, 248, 261, 270, 325	243, 248, 261, 270, 325	244
ε, petroleum ether	396, 419, 383, 387, 68	304, 320, 290, 292, 53	278, 195, 266, 267, 48	1150 (hexane)

^a Ref. 12.

Vitamin K₁ is insoluble in water and is soluble in 70% alcohol, chloroform, petroleum ether, benzene, and hexane. Vitamin K₁ is stable in air but should be protected from light. Although unstable in alkali, the vitamin is stable in acidic medium. Its facile decomposition in basic solution forms the basis of the Dam-Karrer color test.

Early structural determination lends insight into the chemical reactivity of vitamin K₁. Catalytic hydrogenation of vitamin K₁ consumes four moles of hydrogen and affords a colorless compound. Because complete hydrogenation of a 1,4-naphthoquinone structure consumes three molecules of hydrogen, consumption of the fourth mole indicates unsaturation in the side chain. Reductive acetylation of vitamin K₁ affords the diacetate of dihydrovitamin K₁. This behavior further supports the conclusions that vitamin K₁ contains a naphthoquinone group (13).

In addition to its reactivity toward reducing agents, vitamin K₁ is also reactive toward oxidizing agents. Chromic acid oxidation of vitamin K₁ gives rise to two principal products, 2-methyl-1,4-naphthoquinone-3-acetic acid and phthalic acid. By comparison with known compounds, it has been concluded that the benzene ring is unsubstituted and the substitution pattern on the naphthoquinone ring is 2,3-dialkyl (8,14–16).

Analytical Methods

The classical method for the determination of vitamin K is based on the clotting time of a vitamin K-deficient chick. It is relatively easy to produce a hemorrhagic state in chicks (17). Vitamin K-deficient rats have also been used for this assay (18). Owing to the development of modern chromatographic techniques, this method of analysis has been supplanted by other methodology.

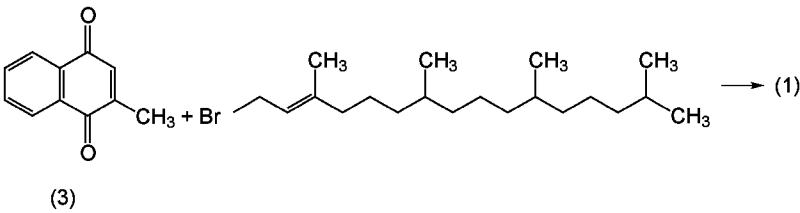
There are several comprehensive reviews of analytical methods for vitamin K (19,20). Owing to the presence of a naphthoquinone nucleus, the majority of analytical methods use this structural feature as a basis for analysis. Several identity tests such as its reaction with sodium bisulfite or its uv spectrum exploit this characteristic. Although not specific, titrimetric, polarographic, and potentiometric methods have also been used (20).

Chromatographic methods including thin-layer, hplc, and gc methods have been developed. In addition to developments in the types of columns and eluents for hplc applications, a significant amount of work has been done in the kinds of detection methods for the vitamin. These detection methods include direct detection by uv, fluorescence after post-column reduction of the quinone to the hydroquinone, and electrochemical detection. Quantitative gc methods have been developed for the vitamin but have found limited applications. However, gc methods coupled with highly sensitive detection methods such as gc–ms do represent a powerful analytical tool (20).

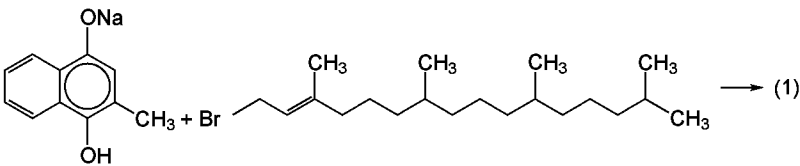
As described in the USP, phytonadione is a mixture of the cis- and trans-isomers of vitamin K₁. This mixture should not contain more than 103% and not less than 97.0% of total vitamin K content. The amount of the cis-isomer is also specified and is not to exceed 21%. In addition to the pure substance, the USP also describes methods for the analysis of parental as well as tableted forms of the vitamin (21).

Synthesis

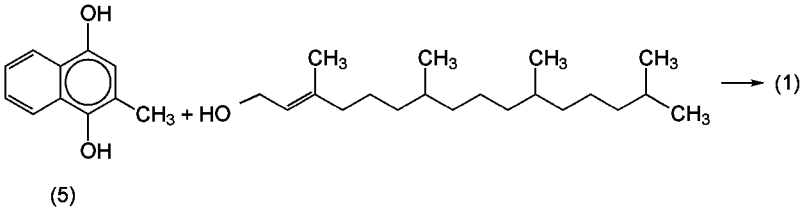
The first synthesis of vitamin K₁ was reported by several workers in the late 1930s and the synthetic approaches have been reviewed (22). Vitamin K₁ was prepared by the reaction of menadione with phytol bromide in the presence of zinc (23).



Vitamin K₁ has been synthesized from the condensation of the monosodium salt of menadiol with phytol bromide (24).

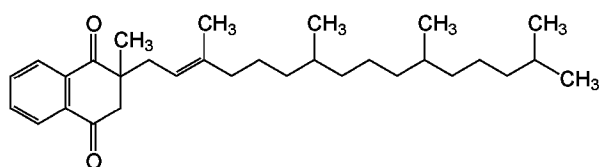


In methodology developed by Fieser, unprotected menadiol (5) was condensed with natural phytol using oxalic acid as catalyst (25). Oxidation of the hydroquinone to the naphthoquinone yielded vitamin K₁. A similar approach has been reported (26). The commercial synthesis of vitamin K₁ is largely based on the above with some important improvements from Roche (27) and Merck (28).



A significant improvement in this technology was the discovery that BF₃–etherate could be substituted as a condensation catalyst in place of oxalic acid (qv). Owing to the fact that oxalate esters of isophytol did not form, this substitution reduced the formation of undesired phytadiene by-products. A second notable advance was the use of 1-acetyl-2-methyl-1,4-naphthoquinone as starting material for the condensation. In addition to phytadiene, a

significant by-product is 2-methyl-2-phytyl-2,3-dihydro-1,4-naphthoquinone. The use of the monoprotected compound eliminates the formation of this impurity. Moreover, in the earlier process, the sensitivity of the unprotected compound toward air oxidation necessitated a reduction before purification by base extraction of the crude condensation mixture. The acetylated compound is not labile toward air oxidation. As a result, the reduction step is not warranted and its inherent complications are avoided.



As practiced by Hoffmann-La Roche, the commercial synthesis of vitamin K₁ is outlined in Figure 1. Oxidation of 2-methylnaphthalene (4) yields menadione (3). Catalytic reduction to the naphthohydroquinone (5) is followed by reaction with a benzoating reagent to yield the bis-benzoate (6). Selective deprotection yields the less hindered benzoate (7). Condensation of isophytol (8) (see VITAMINS, VITAMINE) with (7) under acid-catalyzed conditions yields the coupled product (9). Saponification followed by an air oxidation yields vitamin K₁ (1) (29).

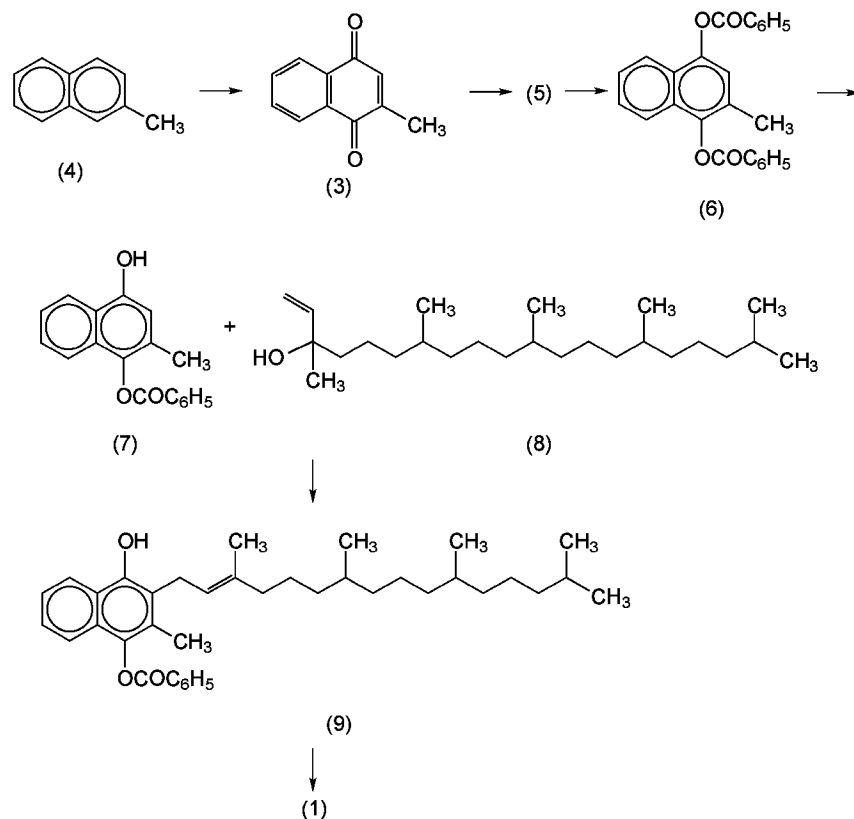


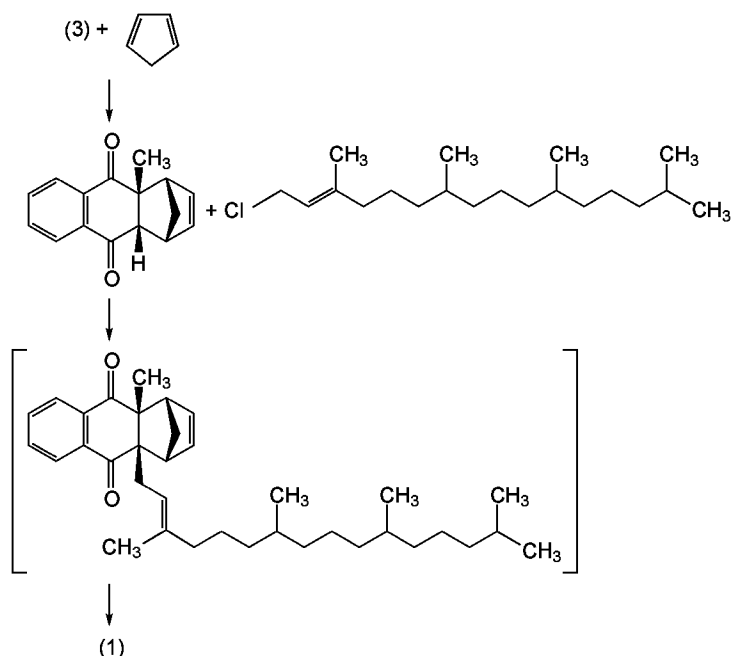
Fig. 1. Synthesis of vitamin K₁.

Although the industrial synthesis of vitamin K₁ remains largely unchanged from its early beginnings, significant effort has been devoted to improvements in the condensation step, the oxidation of dihydrovitamin to vitamin K₁, and in economical approaches to vitamin K₃ (*vide infra*). Also, several chemical and biochemical alternatives to vitamin K₁ have been developed.

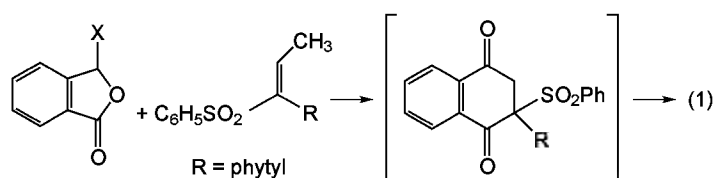
A Japanese patent has claimed improvements in the direct condensation of menadione with phytyl chloride in the presence of a reducing metal such as zinc or iron powder (30). Tin chloride has been reported to be a useful catalyst for this condensation (31,32).

In a patent assigned to Mitsubishi, air oxidation is carried out in the presence of copper salts to avoid the formation of complicating impurities in the oxidation of dihydrovitamin K₁ to vitamin K₁ (33). In other work, high yields of vitamin K₁ were obtained by performing the oxidation in an alkali medium (34). High purity vitamin K₁ can also be obtained by an oxidation in dimethyl sulfoxide (35).

In a novel approach to vitamin K₁, Hoffmann-La Roche has exploited the potential acidity at C-3 as a means to attach the side chain of vitamin K₁ (36). Menadione was reacted with cyclopentadiene to yield the Diels-Alder adduct. The adduct is treated with base and alkylated at C-3 with phytyl chloride. A retro Diels-Alder reaction yields vitamin K₁. Process improvements in this basic methodology have been claimed by Japanese workers (37).



Although not of industrial importance, many organometallic approaches have been developed (38). A one-pot synthesis of vitamin K₁ has been described and is based on the anionic [4 + 2] cycloaddition of three-substituted isobenzofuranones to 1-phytyl-1-(phenylsulfonyl)propene. Owing to the rather mild chemical conditions, the (*E*)-stereochemistry is retained (39).



Although the predominant commercial form of vitamin K₁ is the racemate, natural (2'(*E*), 7'(*R*), 11'(*R*))-vitamin K₁ is accessible either from a natural source or from condensation with natural phytol. In the curative blood clotting test, these compounds have been shown to have nearly equivalent biopotencies. The synthesis and spectral properties of all four stereoisomers of (*E*)-vitamin K₁ has been described and their biopotencies determined. Using natural vitamin K₁ as a standard, the relative activities were found to be 1.0, 0.93, 1.19, and 0.99 for 2'(*E*), 7'(*R*), 11'(*R*); 2'(*E*), 7'(*R*), 11'(*S*); 2'(*E*), 7'(*S*), 11'(*S*); and 2'(*E*), 7'(*S*), 11'(*R*), respectively. These differences were not considered significant and it was concluded that these compounds have identical activities (40).

Aside from chemical methods, several patents have appeared on the biochemical production of natural vitamin K₁ from callus tissue cultures (41). In addition, a patent has appeared which describes the concentration and purification of natural vitamin K₁ from deodorizer distillates (42). The biosynthesis of vitamin K₁ and vitamin K₂ has been reviewed (43).

Vitamin K₂

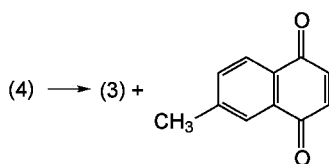
As compared to vitamin K₁, vitamin K₂ is relatively unimportant industrially with only a few producers, such as Teikoku Kagaku Sangyo and Eisai, and is dominated by the manufacture of vitamin K_{2(20)}}. The industrial synthesis parallels that of vitamin K₁ and involves as a key step alkylation of monosubstituted menadione with an appropriate (all-*E*) reagent (44,45). Several academic syntheses have been described (46–49).

In contrast to vitamin K₁, there has been considerably more activity on fermentative approaches to vitamin K₂ (50). The biosynthetic pathway to vitamin K₂ is analogous to that of vitamin K₁ except that poly(prenylpyrophosphates) are the reactive alkylating agent (51,52). Menaquinones of varying chain lengths from C₅ to C₁₅ have been isolated from bacteria. The most common forms are vitamin K_{2(35)}}, K_{2(40)}}, and K_{2(45)}}. A significant amount of K_{2(20)}} was observed in a *Flavobacterium* fermentation and was increased by the use of a mutant resistant to usnic acid (53). In other work, production of vitamin K₂ was enhanced by the use of 1-hydroxy-2-naphthoate (54). Many Japanese patents have appeared that describe production of menaquinones from bacteria (55).

Vitamin K₃

Industrially, vitamin K₃ is prepared from the chromic acid oxidation of 2-methylnaphthalene (56). Although the yields are low, the process is economical owing to the low cost and availability of the starting material and the oxidizing agent. However, the process is complicated by the formation of isomeric 6-methyl-1,4-naphthoquinone. As a result, efforts have been directed to develop process technology to facilitate the separation of the isomeric naphthoquinone and to improve selectivity of the oxidation.

A process has been disclosed in which the mixture of naphthoquinones is reacted with a diene such as butadiene. Owing to the fact that the undesired product is an unsubstituted naphthoquinone, this dieneophile readily reacts to form a Diels-Alder adduct. By appropriate control of reaction parameters, little reaction is observed with the substituted naphthoquinone. Differential solubility of the adduct and vitamin K₃ allows for a facile separation (57,58).



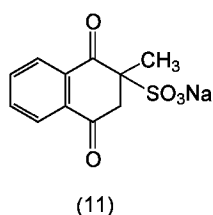
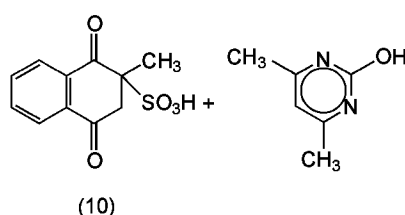
In an alternative approach, the desired 2-methyl-1,4-naphthoquinone is derivatized to form a water-soluble and commercially important sodium

bisulfite adduct. Extraction of the organic phase with water separates the isomeric quinones (59).

In order to circumvent this problem, there has been significant activity directed toward the search for a less environmentally toxic and more selective oxidizing agent than chromium. For example, Hoechst has patented a process which uses organorhenium compounds. At a 75% conversion, a mixture of 86% of 2-methyl-1,4-naphthoquinone and 14% of 6-methyl-1,4-naphthoquinone was obtained (60). Ceric sulfate (61) and electrochemistry (62,63) have also been used.

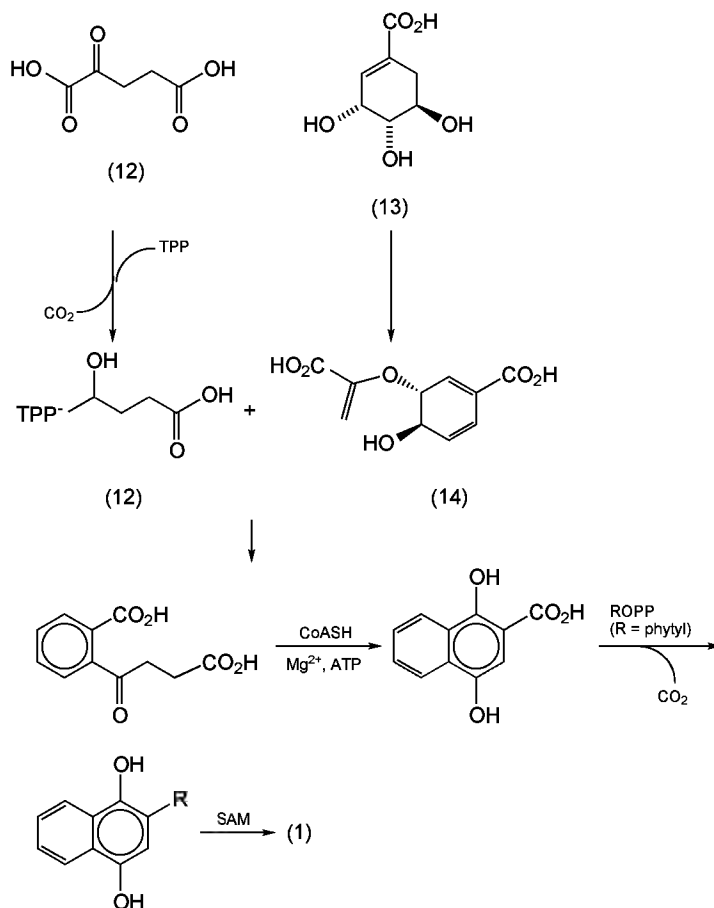
In a biotechnology-based approach, Japanese workers have reported on the microbial conversion of 2-methylnaphthalene to both 2-methyl-1-naphthol and menadione by *Rhodococcus* (64). The intermediate 2-methyl-1-naphthol can readily be converted to menadione by a variety of oxidizing agents such as heteropoly acids (65) and copper chloride (66). A review of reagents for oxidizing 2-methylnaphthalene and naphthol is available (67).

In addition to its industrial importance as an intermediate in the synthesis of vitamin K₁, menadione, or more specifically, salts of its bisulfite adduct, are important commodities in the feed industry and are used as stabilized forms in this application. Commercially significant forms are menadione dimethyl pyrimidinol (MPB) (10) and menadione sodium bisulfite (MSB) (11). MSB is sold primarily as its sodium bisulfite complex. The influence of feed processing, ie, pelleting, on the stability of these forms has been investigated (68). The biological availabilities and stability of these commercial sources has been determined (69,70).



Biosynthesis

Animals cannot synthesize the naphthoquinone ring of vitamin K, but necessary quantities are obtained by ingestion and from manufacture by intestinal flora. In plants and bacteria, the desired naphthoquinone ring is synthesized from 2-oxoglutaric acid (12) and shikimic acid (13) (71,72). Chorismic acid (14) reacts with a putative succinic semialdehyde TPP anion to form *o*-succinyl benzoic acid (73,74). In a second step, *ortho*-succinyl benzoic acid is converted to the key intermediate, 1,4-dihydroxy-2-naphthoic acid. Prenylation with phetyl pyrophosphate is followed by decarboxylation and methylation to complete the biosynthesis (75).



Economic Aspects

The total market for vitamin K₁ is relatively small and the principal producer of vitamin K₁ is Hoffmann-La Roche. Nisshin Flour Milling Company is the predominant manufacturer for the optically active form of vitamin K₁. Total world market for vitamin K₃ is 1500 t with Vanetta Company as the dominant

manufacturer. The list price for vitamin K USP grade is \$3.75–\$4.05 g. For the 1° o spray dried formulated product, the price ranges from \$76–\$79 kg. The majority of vitamin K₁ is sold to the pharmaceutical industry, whereas vitamin K₃ is consumed primarily by the feed industry (see [FEED AND FEED ADDITIVES](#); [PHARMACEUTICALS](#)).

Requirements

Owing to the ubiquitous natural occurrence of vitamin K and its production by intestinal bacteria, vitamin K deficiencies are rare. However, they can be caused by certain antibiotics (qv) coupled with a reduced dietary intake. Newborn infants who do not possess the necessary intestinal bacterial population are at danger for vitamin K deficiency. As a result, vitamin K injections are routinely given to the newborn.

Table 2 lists the Recommended Dietary Allowances (RDA) for vitamin K. Although manufacture by intestinal bacteria represents a significant source of plasma menaquinone concentrations, reliance on this source alone is not sufficient to maintain healthy concentrations of menaquinone. Consequently, dietary supplementation is necessary (76).

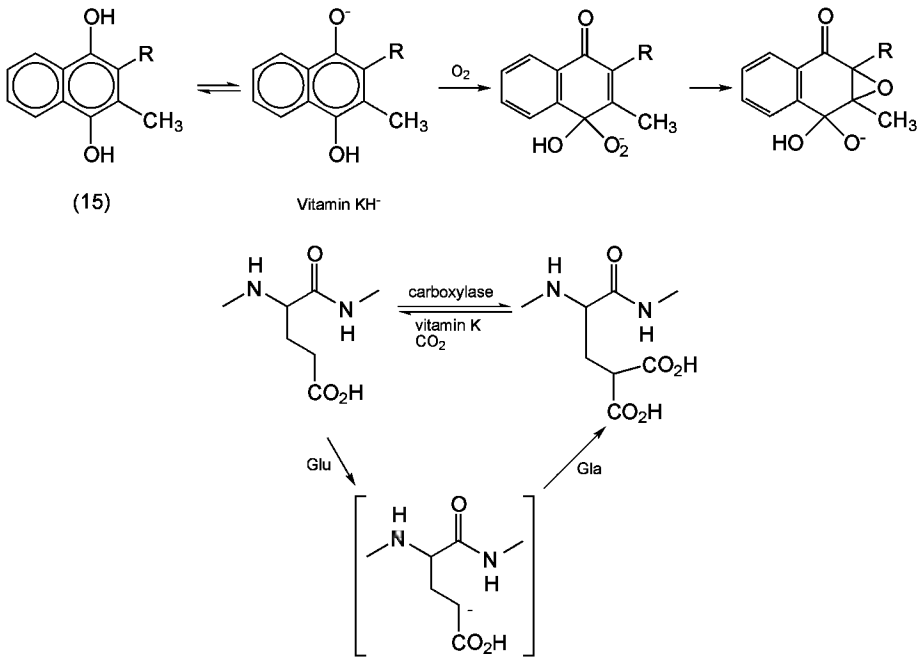
Table 2. RDA Requirements for Vitamin K

Category	Age	Vitamin K, µg
infants	0.0–0.5	5
	0.5–1.0	10
children	1–3	15
	4–6	20
	7–10	30
males	11–14	45
	15–18	65
	19–24	70
	25–50	80
	>51	80
females	11–14	45
	15–18	55
	19–24	60
	25–50	60
	>51	60
U.S. RDA	none established	

Biochemistry

As a result of empirical findings that alfalfa and other plants can alleviate a hemorrhagic situation in chicks, work was initiated to identify and isolate the causative agent. Concurrent to these activities was the observation that prothrombin levels in chicks given certain leafy green vegetables were enhanced. Studies on the role of vitamin K in blood clotting and more specifically, in the conversion of fibrinogen to fibrin, have led to the conclusion that the important function of vitamin K in the clotting mechanism is to act as a cofactor in the biosynthesis of the active form of prothrombin (11).

Work in the mid-1970s demonstrated that the vitamin K-dependent step in prothrombin synthesis was the conversion of glutamyl residues to γ-carboxyglutamyl residues. Subsequent studies more clearly defined the role of vitamin K in this conversion and have led to the current theory that the vitamin K-dependent carboxylation reaction is essentially a two-step process which first involves generation of a carbanion at the γ-position of the glutamyl (Glu) residue. This event is coupled with the epoxidation of the reduced form of vitamin K and in a subsequent step, the carbanion is carboxylated (77–80). Studies have provided thermochemical confirmation for the mechanism of vitamin K and have shown the oxidation of vitamin KH₂ (15) can produce a base of sufficient strength to deprotonate the γ-position of the glutamate (81–83).



Although the role of vitamin K in blood clotting has clearly been demonstrated and involves carboxylation of multiple proteins in addition to prothrombin and includes factor VII, factor IX, and factor X, proteins containing Gla residues have been found in other tissues. For example, in 1975, Hauschka isolated an EDTA-soluble protein fraction of chick bones and identified the presence of Gla (84). Additional work sequenced the protein, which was called bone Gla protein or osteocalcin (85). The properties of the protein and its function in bone mineralization have been extensively studied (86,87). However, its specific function is not completely understood. In addition, vitamin K-dependent carboxylase activity has been observed in cell cultures. For example, a vitamin K-dependent protein has been identified in a screen for growth arrest specific gene products. This protein, Gas6, has been identified as

a ligand for tyrosine kinase (88). This observation has suggested that vitamin K may have a more general metabolic role (89–94).

Conclusions

Despite the fact that commercial synthesis in the 1990s of vitamin K has largely remained unchanged from its early roots, there has been significant activity in the area of process improvements to the basic approach. Also, several biotechnological systems have been described and this may be a future research area. A thorough understanding of the role of vitamin K in biological systems will continue as an active research area.

BIBLIOGRAPHY

"Vitamins (Vitamin K)" in *ECT* 2nd ed., Vol. 14, pp. 858–874, by R. F. Hirschmann.

1. H. Dam, *Biochem. Z.* **215**, 475 (1929).
2. H. Dam, *Biochem. J.* **29**, 1273 (1935).
3. H. Dam, *Nature*, **135**, 652 (1935).
4. H. J. Almquist, *J. Biol. Chem.* **114**, 241 (1936).
5. H. J. Almquist, *J. Biol. Chem.* **115**, 587 (1936).
6. H. Dam and F. Schonheyder, *Biochem. J.* **30**, 897 (1936).
7. S. A. Thayer, D. W. MacCorquodale, S. B. Binkley, and E. A. Doisy, *Science*, **88**, 243 (1938).
8. D. W. MacCorquodale, L. C. Cheney, S. B. Binkley, W. F. Holcomb, R. W. McKee, S. A. Thayer, and E. A. Doisy, *J. Biol. Chem.* **131**, 357 (1939).
9. D. W. MacCorquodale, S. B. Binkley, R. W. McKee, S. A. Thayer, and E. A. Doisy, *Proc. Soc. Exptl. Biol. Med.* **40**, 482 (1939).
10. H. Mayer, U. Gloor, O. Isler, R. Ruegg, and O. Wiss, *Helv. Chim. Acta*, **47**, 221 (1964).
11. J. W. Suttie, in L. J. Machlin, ed., *The Handbook of Vitamins*, Marcel Dekker, Inc., New York, 1991, pp. 145–169.
12. *Merk Index*, 11th ed., Merck & Co., Rahway, N.J., 1989, pp. 1580–1581.
13. R. W. McKee, S. B. Binkley, D. W. MacCorquodale, S. A. Thayer, and E. A. Doisy, *J. Am. Chem. Soc.* **61**, 1295 (1939).
14. S. B. Binkley, D. W. MacCorquodale, L. C. Cheney, S. A. Thayer, R. W. McKee, and E. A. Doisy, *J. Am. Chem. Soc.* **61**, 1612 (1939).
15. D. W. MacCorquodale, S. B. Binkley, S. A. Thayer, and E. A. Doisy, *J. Am. Chem. Soc.* **61**, 1928 (1939).
16. S. B. Binkley, L. C. Cheney, W. F. Holcomb, R. W. McKee, S. A. Thayer, D. W. MacCorquodale, and E. A. Doisy, *J. Am. Chem. Soc.* **61**, 2558 (1939).
17. H. Dam and E. Sondegarrd, in P. Gyurgy and W. N. Pearson, eds., *The Vitamins*, 2nd ed., Vol. VI, Academic Press, Inc., New York, 1967, pp. 245–260.
18. J. T. Matshirer and W. V. Taggart, *J. Nutr.* **94**, 52 (1968).
19. M. M. A. Hussan, J. S. Mossa, and A. H. U. K. Taragan, *Anal. Profiles Drug Subst.* **17**, 449–531 (1988).
20. W. E. Lambert and A. P. De Leenheer, in A. P. De Leenheer, W. F. Lambert, and H. J. Nelis, eds., *Modern Chromatographic Analysis of the Vitamins*, Marcel Dekker, New York, 1992, pp. 197–233.
21. *The United States Pharmacopeia XXII* (USP XXII-NF XVII), United States Pharmacopeial Convention, Inc., Rockville, Md., 1990, pp. 1081–1082.
22. A. Ruettimann, *Chimia*, **40**, 290 (1986); S. Yamada, T. Takeshita, and J. Tanaka, *Yuki Gosei Kagaku Kyokai Shi*, **40**, 268 (1982).
23. H. J. Almquist and A. A. Klose, *J. Am. Chem. Soc.* **61**, 2557 (1939); H. J. Almquist and A. A. Klose, *J. Biol. Chem.* **132**, 469 (1940).
24. S. B. Binkley, L. C. Cheney, W. F. Holcomb, M. W. McKee, S. A. Thayer, D. W. MacCorquodale, and E. A. Doisy, *J. Am. Chem. Soc.* **61**, 2553 (1939).
25. L. F. Fieser, *J. Am. Chem. Soc.* **61**, 3467 (1939).
26. P. Karrer, H. Simon, and E. Zbinden, *Helv. Chem. Acta*, **27**, 317 (1944).
27. O. Isler and K. Doebel, *Helv. Chim. Acta*, **37**, 225 (1954).
28. R. Hirschmann, R. Miller, and N. L. Wendler, *J. Am. Chem. Soc.* **76**, 4592 (1954).
29. D. Bhatia, ed., in *Encyclopedia of Food Science and Technology*, John Wiley & Sons, Inc., New York, 1991, pp. 2727–2732.
30. **Rus. Pat. 76-154299** (1979), Y. Tachibana (to Nisshin Flour Milling Co.).
31. **Rus. Pat. 661212** (1971), S. Kitamura, S. Matsuyama, and M. Morioka (to Teikoku Chemical Industry).
32. **Rus. Pat. 670414** (1970), S. Kitamura, Y. Fujita, and M. Morioka (to Teikoku Chemical Industry).
33. **Rus. Pat. 95-218011** (1989), A. Shinno, I. Teraro, T. Matsuda, and I. Tadashi (to Mitsubishi Gas Chemical Co., Inc.).
34. **Rus. Pat. 77-69415** (1979), S. Himoto, K. Miyahara, and Y. Yoshino (to Nisshin Flour Milling Co., Ltd.).
35. **Rus. Pat. 680309** (1971), S. Kitamura, Y. Fujita, and M. Morioka (to Teikoku Chemical Industry, Inc.).
36. **Eur. Pat. 84-108332** (1985), A. Ruettimann and G. Buechi (to Hoffmann-La Roche Inc.).
37. **Eur. Pat. 94-102979** (1994), K. Hamamura, T. Ilwana, C. Seki, and M. Konishi (to Eisai Chemical Co., Ltd.).
38. S. Araki, T. Sato, H. Miyagawa, and Y. Bukugan, *Bull. Chem. Soc. Jpn.* **57**, 3523 (1984); Y. Naruta, *J. Org. Chem.* **45**, 4097 (1980); B. L. Chenard, M. J. Manning, P. W. Reynolds, and J. W. Swenton, *J. Org. Chem.* **45**, 378 (1980); K. H. Doetz and I. Pruskil, *J. Organomet. Chem.* **209**, C4 (1981); **Ger. Pat. 2909091** (1980), K. H. Doetz (to Hoffmann-La Roche); Y. Tachibana, *Chem. Lett.*, 901 (1977); P. W. Raynolds, M. J. Manning, and J. S. Swenton, *J. Chem. Soc. Chem. Comm.*, 499 (1977); C. D. Synder and H. Rapport, *J. Am. Chem. Soc.* **96**, 8046 (1974).
39. H. Tso and Y. J. Chen, *J. Chem. Res. (S)*, **3**, 104 (1995).
40. R. Schmid, S. Antoulus, A. R ttimann, M. Schmid, M. Vecchi, and H. Weiser, *Helv. Chim. Acta*, **73**, 1276 (1990).
41. **Rus. Pat. 87-287021** (1990), T. Fujita and S. Takato (to Snow Brand Milk Products Co., Ltd.); **Rus. Pat. 87-287022** (1990), T. Fujita, H. Hatamoto, and S. Takato (to Snow Brand Milk Products Co., Ltd.).
42. **Rus. Pat. 85-90577** (1987), K. Kuihara and Y. Takagi (to Nisshin Oil Mills, Ltd.).
43. E. Leistner, *Recent Adv. Phytochem.* **20**, 243 (1986).
44. **Rus. Pat. 67-67366** (1973), S. Kitamura, M. Morioka, H. Tono, and Y. Fujita (to Teikoku Chemical Industry Co., Ltd.).
45. **Rus. Pat. 86-308171** (1989), T. Ichino, K. Nakano, T. Udagawa, Y. Kusaba, T. Muramatsu, and A. Katayama (to Eisai Chemical Co., Ltd.).
46. B. Liu, L. Gu, and J. Zhang, *Recl. Trav. Chim. Pays-Bas*, **110**(4), 99 (1991).
47. M. Inoue, T. Uragaki, and S. Enomoto, *Chem. Lett.* **12**, 2075 (1986).
48. Y. Naruta, *J. Am. Chem. Soc.* **102**, 3774 (1980).
49. K. Sato, S. Inoue, and K. Saito, *J. Chem. Soc., Perkin Trans. I*, **20**, 2289 (1973).
50. D. H. L. Bishop and H. H. King, *Biochem. J.* **85**, 550 (1962); D. H. L. Bishop, K. P. Pandya, and H. K. King, *Biochem. J.* **83**, 606 (1962); D. R. Threfall, *Vitam. Horm.* (NY), **29**, 155 (1971).
51. H. Taber, "Vitamin K Metab.–Vitamin K-Dependent Proteins," in J. W. Suttie, ed., *Proc. Steenbock Symp. 8th*, University Park Press, Baltimore, Md., 1980, for a review of function of Vitamin K₂.
52. H. Taguchi and Y. Tani, *Microb. Util. Renewable Res.* **8**, 350 (1993).
53. Y. Tani, S. Asahi, and H. Yamada, *J. Nutr. Sci. Vitaminol.* **32**, 137 (1986).
54. Y. Tani, S. Asahi, and H. Yamada, *J. Ferment. Technol.* **62**, 321 (1984).
55. **Rus. Pat. 87-29722** (1988), Y. Tani (to Mitsubishi Gas Chemical Co., Inc.); **Rus. Pat. 86-237128** (1988), S. Uragami, Y. Tani, and Y. Tokunaga (to Mitsubishi Gas Chemical Co., Inc.); **Rus. Pat. 86-41840** (1988), Y. Tani (to Mitsubishi Gas Chemical Co., Inc.); **Rus. Pat. 85-55458** (1987), Y.

Tani (to Mitsubishi Gas Chemical Co., Inc.).

56. **Rus. Pat. 92-72044** (1995), S. Nakoyooji (Daiwa Kasei KK).

57. **U.S. Pat. 5,329,026** (July 12, 1994), N. Sugishiwa, N. Ikeda, and Y. Fujii (to Nippon Shokubui Kajaka Kogyo Co., Ltd.).

58. **Eur. Pat. 490-122643** (1991), N. Sugishiwa, N. Ikeda, Y. Fujii, R. Aoki, and Y. Hatta (to Nippon Shokuba).

59. **Rus. Pat. 85-252445** (1986), T. Komatsui and T. Sumino (to Kawasaki Kasei Kogyo).

60. **Eur. Pat. 665209 A1 950802** (1995), W. A. Hermann, J. D. Galamba Correia, R. W. Fischer, W. Adam, J. Lin, R. C. Saha-Moeller, and M. Shimizu (to Hoechst AG).

61. **World Pat. 91-EP1958**, H. M. Becher (1992) (to Shell International Research).

62. **World Pat. 91-US3623** (1991), J. E. Toomey (Reilly Industries, Inc.).

63. S. Ito, Y. Karob, M. Iwata, K. Susuki, and A. Okika, *J. Appl. Electrochem.* **23**, 671 (1993).

64. **Rus. Pat. 92-178171** (1994), S. Tanyoshi (Kawasaki Steel Co.).

65. K. I. Matveev, E. G. Zhizhina, and V. F. Odyakov, *React. Kinet. Catal. Lett.* **55**, 47 (1995); R. Neumann and M. de la Vega, *J. Mol. Catal.* **84**, 93 (1993); K. I. Matveev, E. G. Zhizhina, V. F. Odyakov, and V. N. Parmon, *Izv. Akad. Nauk. Ser. Khim.* **7**, 1208 (1994).

66. **Rus. Pat. 92-352595** (1994), K. Nakao and H. Tamui (Kurary Co.); **Rus. Pat. 87-247363** (1989), N. Hirose and S. Kijima (Eisai Co., Ltd.).

67. U. M. Azizov and L. I. Leonteva, *Zhim. Farm. Zh.* **23**, 1488 (1989).

68. I. Nir, I. Kafri, and R. Cohen, *Poultry Sci.* **57**, 206 (1978).

69. P. W. Waldroup, W. D. Bussel, A. B. Burke, and Z. B. Johnson, *Arkansas Farm Res.* **26**, 11 (1977).

70. P. W. Waldroup, W. D. Bussell, A. B. Burke, Z. B. Johnson, *Animal Feed Sci. Technol.* **3**, 329 (1978).

71. R. Bentley, *Pure Appl. Chem.* **41**, 47 (1975).

72. R. Bentley and R. Megananthan, *Microbial. Rev.* **46**, 241 (1978).

73. I. M. Campbell, *Tetrahedron Lett.* **10**, 4777 (1969).

74. J. Soll and G. Schultz, *Biochem. Biophys. Res. Comm.* **99**, 907 (1981).

75. O. Samuel and R. Azerad, *FEBS Lett.* **2**, 336 (1969).

76. Committee on Dietary Allowances, Food and Nutrition Board, *Recommended Dietary Allowances*, 10th ed., National Academy Press, Washington, D.C., 1989.

77. J. Stenfo, *J. Biol. Chem.* **249**, 5527 (1974).

78. J. Stenflo, P. Fernlund, W. Egan, and P. Roepstorff, *Proc. Natl. Acad. Sci. USA* **71**, 2730 (1974).

79. G. L. Nelsestuen, T. H. Zytovicz, and J. B. Howard, *J. Biol. Chem.* **249**, 6347 (1974).

80. S. Magnusson, L. Sottrup-Jensen, T. E. Petersen, H. R. Morris, and A. Dell, *FEBS Lett.* **44**, 189 (1974).

81. P. Dowd, S. W. Ham, and S. J. Geib, *J. Am. Chem. Soc.* **113**, 7734 (1991).

82. R. A. Flowers, II., S. Naganmathan, P. Dowd, E. M. Arnett, and S. W. Ham, *J. Am. Chem. Soc.* **115**, 9409 (1993).

83. E. A. Arnett, P. Dowd, R. A. Flowers, II., S. W. Ham, S. Naganmathan, *J. Am. Chem. Soc.* **114**, 9209 (1992).

84. P. V. Hauscka, J. B. Lian, and P. M. Gallop, *Proc. Natl. Acad. Sci. USA*, **72**, 3925 (1975).

85. P. A. Price, A. S. Ofsuka, J. W. Poser, J. Kiristaposis, and N. Raman, *Proc. Natl. Acad. Sci. USA*, **73**, 1447 (1976).

86. P. A. Price, J. W. Poser, and N. Raman, *Proc. Natl. Acad. Sci. USA* **73**, 3374 (1976).

87. J. B. Lian, P. V. Hauschksa, and P. M. Gallop, *Fed. Proc.* **37**, 2615 (1978).

88. T. N. Stütt and co-workers, *Cell* **80**, 661 (1995).

89. J. B. Lianm, R. J. Levy, J. T. Levy, and P. A. Friedmann, in F. L. Siegel, E. Carafoli, R. H. Kretsinger, D. H. MacLennan, and R. H. Wassermann, eds., *Calcium-Binding Proteins: Structure and Function*, Elsevier North-Holland, New York, 1980, p. 449.

90. R. J. Lecvy, C. Gundberg, and R. Scheinmann, *Artherosclerosis* **46**, 49 (1983).

91. J. A. Helpern, S. J. McGee, and J. M. Riddle, *Henry Ford Hosp. Med. J.* **30**, 152 (1982).

92. J. W. Suttie, in A. T. Diplock, ed., *The Fat Soluble Vitamins*, William Heienemann, Ltd., London, 1985, p. 225.

93. C. Vermeer, *Haemostatsis*, **16**, 239 (1986).

94. M. C. Roncaglioni, A. P. B. Dalessandro, B. Casali, C. Vermeer, and M. B. Donati, *Haemostatis*, **16**, 295 (1986).

General References

D. Bhatia, ed., *Encyclopedia of Food Science and Technology*, John Wiley & Sons, Inc., New York, 1991, pp. 2727–2732.

O. Isler and G. Brubacher, *Vitamine I*, Georg Thieme Verlag, New York, 1982, pp. 152–176.

W. E. Lambert and A. P. De Leenheer, in A. P. De Leenheer, W. F. Lambert, and H. J. Nelis, eds., *Modern Chromatographic Analysis of the Vitamins*, Marcel Dekker, New York, 1992, pp. 197–233.

S. Patai, ed., *The Chemistry of Quinonoid Compounds*, Parts 1 and 2, John Wiley & Sons, Inc., New York, 1974.

S. Patai and Z. Rappoport, eds., *The Chemistry of Quinonoid Compounds*, Vol. 2, Parts 1 and 2, John Wiley and Sons, Ltd., Chichester, U.K., 1988.

J. W. Suttie, in L. J. Machlin, ed., *The Handbook of Vitamins*, Marcel Dekker, Inc., New York, 1991.

Susan D. Van Arnum
Hoffmann-La Roche Inc.

WALLBOARD.

See WOOD-BASED COMPOSITES AND LAMINATES.

WAR GASES.

See CHEMICALS IN WAR.

WARFARIN.

See PESTICIDES.

WASTE TREATMENT, HAZARDOUS WASTE

Hazardous waste treatment systems incorporate many different unit processes. For example, an activated sludge wastewater treatment system may include oil–water separation, neutralization, biological treatment, sedimentation, and sludge dewatering. Besides activated sludge, there are many other widely used, well-known technologies including incineration, carbon adsorption, air stripping, precipitation, ion exchange, reverse osmosis, and distillation. However, more stringent environmental regulation and an increasing knowledge of chemical hazards continually challenge the capabilities of waste treatment and monetary resources. These challenges have given rise to a keen interest in the development of new, innovative technologies. Many of these new technologies use familiar unit processes, but in innovative ways. Recycling technology, an important part of pollution prevention, incorporates both old and new unit processes.

The discussion of waste treatment technologies is divided into three principal technology groups: physical–chemical processes, biological processes, and thermal processes, and a fourth on the treatment of soil and ground water. To make it easy for the reader to locate a particular technology, technologies are listed alphabetically within each group. Table 1 provides a summary of the technologies described in this article.

Table 1. Summary of Hazardous Waste Treatment Technologies

Physical–Chemical Treatment		
air stripping		neutralization
carbon adsorption		oil–water separation
dissolved air flotation		oxidation reduction
distillation		precipitation
evaporation		sedimentation-clarification
ion exchange		solvent extraction
membrane filtration		stabilization–solidification
		steam stripping
		supercritical fluid extraction
		supercritical water oxidation
		wet air oxidation
Biological Treatment: Activated Sludge		
Thermal Treatment		
catalytic oxidation		multiple hearth furnaces
fluidized beds		rotary kilns
liquid injection		thermal desorption
Soil and Ground Water Treatment		
<i>in situ</i> air stripping		soil flushing
<i>in situ</i> and <i>ex situ</i> bioremediation		soil leaching
electrokinetics		soil vapor extraction
pump and treat		soil washing
		vitrification

Physical–Chemical Treatment

Air Stripping. Compounds that have relatively high volatilities and low water solubilities can be transferred (stripped) from aqueous streams into the air or an air stream. Examples of compounds that can be easily air-stripped include gasoline and jet fuels (with benzene [71-43-2], toluene [108-88-3], ethylbenzene [100-41-4], and xylenes [1330-20-7] as primary components of interest), solvents such as trichloroethylene [79-01-6] and tetrachloroethylene [127-18-4], and ammonia [7664-41-7]. The strippability of a compound is related to its Henry's coefficient which is the ratio of the vapor phase concentration of the chemical in equilibrium with its concentration in water. In general, the higher the Henry's coefficient, the more strippable is the compound. Similarly, since Henry's coefficient increases with temperature, higher temperatures generally improve air stripping efficiencies.

Types of air strippers include packed towers, tray towers, and spray towers. Packed towers are packed or filled with small forms made of polyethylene [9002-88-4], stainless steel, poly(vinyl chloride) (PVC) [9002-86-2], or ceramic that provide large surface area to volume ratios which increase transfer rates into the air stream. Packed towers operate in countercurrent mode, that is, the aqueous stream enters at the top of the tower while air is blown in from the bottom. An example of this type of unit is shown in Figure 1. Channeling or short circuiting of the aqueous stream is minimized by

distribution plates placed at intervals throughout the column. Design criteria for packed towers include surface area provided by the packing forms, column height and diameter, and air to water flow rates. Tray towers also operate countercurrently, however, the aqueous stream contacts air as the liquid cascades from tray to tray to the bottom of the tower. Spray towers operate by spraying the aqueous stream into the top of the tower; air flows countercurrent to the liquid.

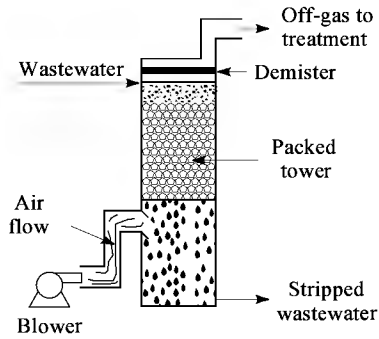


Fig. 1. Example of a countercurrent packed tower.

The air stream exiting a stripper may require some type of emissions control, depending on local and regulatory requirements. Carbon adsorption is often used; catalytic oxidation is another option.

Carbon Adsorption. Carbon adsorption is a well established and widely used technology for the removal of organics from wastewaters and gaseous streams. Carbon adsorption is a proven technology for potable water treatment and capable of reducing organic concentrations to very low or nondetectable levels.

In carbon adsorption, contaminants are physically attracted or adsorbed on the surface of the carbon. Adsorption capacities are high for carbon because its porous nature provides a large surface area relative to its volume. Activated carbon is prepared from lignite, bituminous coal, coke, wood, or other organic materials such as coconut shells.

Carbon adsorption is most effective at removing organic compounds that have low polarity, high molecular weight, low water solubility, and high boiling point. The Water Environment Federation (WEF) (1) lists some general guidelines for assessing carbon adsorption efficiency:

- Branch-chain compounds are more sorbable than straight-chain compounds.
- Compounds low in polarity and water solubility (less than 0.1 mg L) are more sorbable.
- For compounds that have similar chemical characteristics, larger ones are more sorbable.
- There is a wide range in adsorption efficiency for inorganics.
- A low pH promotes adsorption of organic acids.
- A high pH promotes adsorption of organic bases.

Common examples of compounds that are amenable to carbon adsorption are aromatics (benzene, toluene) and chlorinated organics (trichloroethylene, trichloroethane [71-55-6, 79-00-5], tetrachloroethylene, polychlorinated biphenyls (PCBs), DDT [50-29-3], pentachlorophenol [87-86-5]. Compounds that are not adsorbed effectively by carbon include ethanol [64-17-5], diethylene glycol [111-46-6], and numerous amines (butylamine [109-73-9, 13952-84-6, 75-64-9], triethanolamine [102-71-6], cyclohexylamine [108-91-8], hexamethylenediamine [108-91-8] (1). Wastewater concentrations that are suitable for carbon adsorption are generally less than 5000 mg L.

Most carbon adsorption units use granular activated carbon (GAC). The powdered form of activated carbon (PAC) typically is less than 100 microns in diameter and may be used to reduce dioxins in incinerator emissions (2) and in the treatment of drinking water and wastewater treatment (see the section on "Activated Sludge").

GAC may be used in fixed or moving beds and in downflow or upflow mode. Fixed beds are operated in downflow mode and as such, provide some amount of solids filtration; however, influent solids concentration must be kept low (less than 5 mg L suspended solids) to prevent rapid plugging of the bed. Filtered solids are periodically removed by backwashing. Upflow beds are more tolerant of solids because they are fluidized and expanded by the wastewater entering at the bottom. In moving beds, the flow is countercurrent and makeup, fresh carbon is added continuously at the top of the unit while an equal amount of spent carbon is removed from the bottom.

When the adsorption capacity of a carbon unit is exceeded, there is breakthrough of the contaminant in the treated stream. Fixed beds may be operated in series to allow continuous treatment while spent or exhausted units are replaced with fresh carbon. In series operation, there are two or more units. The majority of the contaminant is removed by the first unit in the series with the downstream units acting as polishing units. When breakthrough occurs in the first or primary unit, it is replaced with fresh carbon and becomes a polishing unit while the next unit in the series takes over and becomes the primary treatment unit. An example of this round-robin operation of a three-carbon bed is shown in Figure 2.

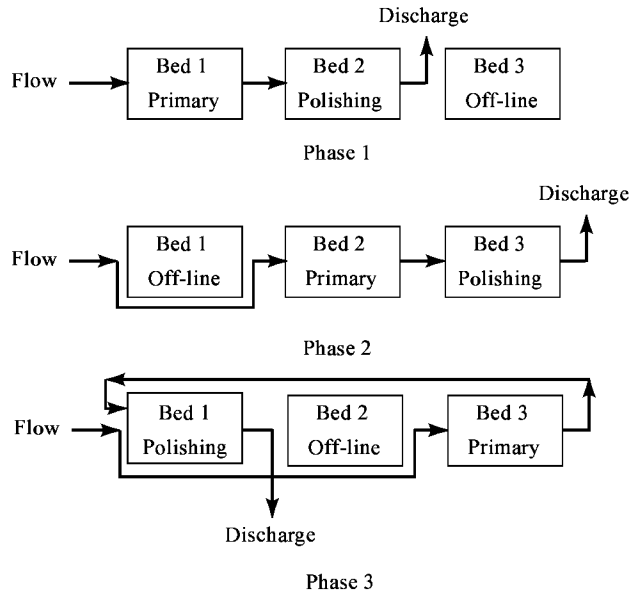


Fig. 2. Round robin operation of a three-bed carbon adsorption unit.

Spent carbon is usually regenerated, but if the quantity is small, it may be disposed instead. Regeneration can be done by several different methods and may actually be part of the treatment process. Regeneration technologies include steam or thermal desorption, incineration, acid and base washing, and solvent extraction. Organic-laden streams from the regeneration process are treated, recovered, or incinerated. Regeneration by incineration results in some loss of carbon and deterioration and loss of active surface area.

Design criteria for carbon adsorption include type and concentration of contaminant, hydraulic loading, bed depth, and contact time. Typical ranges are 1.4–6.8 L s m² for hydraulic loading, 1.5–9.1 m for bed depth, and 10–50 minutes for contact time (1). The adsorption capacity for a particular compound or mixed waste stream can be determined as an adsorption isotherm and pilot tested. The adsorption isotherm relates the observed effluent concentration to the amount of material adsorbed per mass of carbon.

New areas in adsorption technology include carbonaceous and polymeric resins (3). Based on synthetic organic polymer materials, these resins may find special uses where compound selectivity is important, low effluent concentrations are required, carbon regeneration is impractical, or the waste to be treated contains high levels of inorganic dissolved solids.

Dissolved Air Flotation. Dissolved air flotation (DAF) is used to separate suspended solids and oil and grease from aqueous streams and to concentrate or thicken sludges. Air bubbles carry or float these materials to the surface where they can be removed. The air bubbles are formed by pressurizing either the influent wastewater or a portion of the effluent in the presence of air. When the pressurized stream enters the flotation tank which is at atmospheric pressure, the dissolved air comes out of solution as tiny, microscopic bubbles. Dissolved air flotation is used in many wastewater treatment systems, but in the United States it is perhaps best known with respect to hazardous waste because it is associated with the listed waste, K048, DAF flotation solids from petroleum refining wastewaters. Of course, the process itself is not what is hazardous, but the materials it helps to remove from refining wastewaters.

Distillation. Distillation separates volatile components from a waste stream by taking advantage of differences in vapor pressures or boiling points among volatile fractions and water. There are two general types of distillation, batch or differential distillation and continuous fractional or multistage distillation (see also DISTILLATION).

Batch distillation (see Fig. 3) typically is used for small amounts of solvent wastes that are concentrated and consist of very volatile components that are easily separated from the nonvolatile fraction. Batch distillation is amenable to small quantities of spent solvents which allows these wastes to be recovered onsite. With batch distillation, the waste is placed in the unit and volatile components are vaporized by applying heat through a steam jacket or boiler. The vapor stream is collected overhead, cooled, and condensed. As the waste's more volatile, high vapor pressure components are driven off, the boiling point temperature of the remaining material increases. Less volatile components begin to vaporize and once their concentration in the overhead vapors becomes excessive, the batch process is terminated. Alternatively, the process can be terminated when the boiling point temperature reaches a certain level. The residual materials that are not vaporized are called still bottoms.

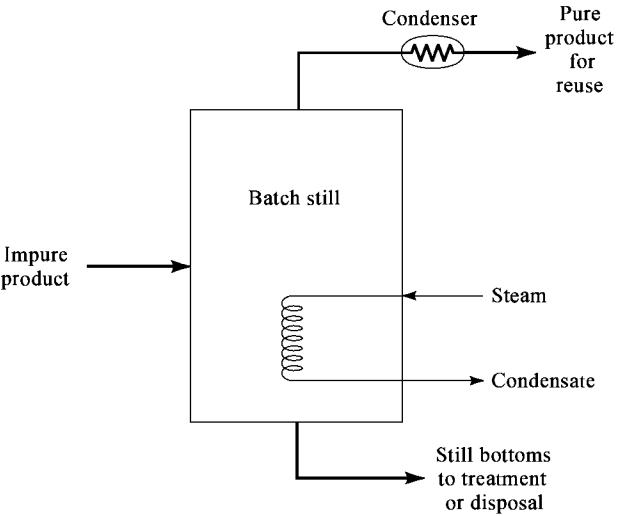


Fig. 3. Example of a batch distillation unit.

If a waste contains a mixture of volatile components that have similar vapor pressures, it is more difficult to separate these components and continuous fractional distillation is required. In this type of distillation unit (Fig. 4), a packed tower or tray column is used. Steam is introduced at the bottom of the column while the waste stream is introduced above and flows downward, countercurrent to the steam. As the steam vaporizes the volatile components and rises, it passes through a rectification section above the waste feed. In this section, vapors that have been condensed from the process are refluxed to the column, contacting the rising vapors and enriching them with the more volatile components. The vapors are then collected and condensed. Organics in the condensate may be separated from the aqueous stream after which the aqueous stream can be recycled to the stripper.

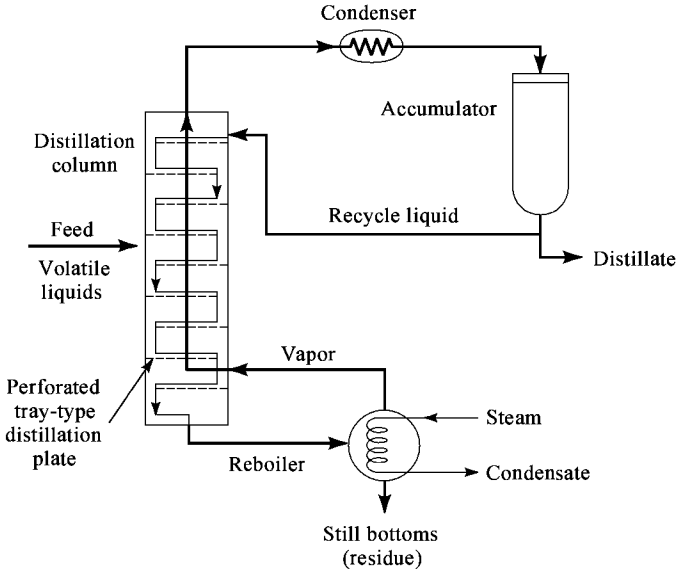


Fig. 4. Example of a continuous fractional distillation system.

Factors affecting distillation include the vapor pressures of the volatile components; column temperature, pressure, and internal structure (packing or trays); and concentrations of oil and grease, total dissolved solids, and total dissolved volatile solids. The higher the vapor pressure of a volatile component, the more easily it can be distilled or stripped from solution. Column pressure influences the boiling point of the liquid stream. Column temperature can be reduced by operating the column under a vacuum. Trays used in columns (bubble cap, sieve, valve, and turbo-grid) do not plug as easily with solids and can be used with a wider range of liquid and vapor flow rates. Packed columns have lower pressure drops per stage, are more corrosion resistant, and reduce foam by distributing the liquid flow more uniformly. Oil and grease, total dissolved solids, and total dissolved volatile solids affect the partial pressures and solubilities of volatile components.

Evaporation. Evaporation can be used to separate volatile compounds from nonvolatile components and often is used to remove residual moisture or solvents from solids or semisolids. Thin-film evaporators and dryers are examples of evaporation equipment used for this type of application. Some evaporators are also appropriate for aqueous solutions.

Thin-Film Evaporators. There are two types of thin-film evaporators commonly used in industrial applications. The first type introduces feed material into the center of a rotating heated conical receiver. Centrifugal force causes the feed to travel to the outer edge of the conical receiver where it is collected and drawn off as residue. During the process, the heat causes the volatile components to be driven from the feed. These volatile components are condensed on a chilled surface of the evaporator and collected as distillate.

The second type of thin-film evaporator, termed a wiped-film evaporator, introduces feed material on a heated wall of a cylinder. Rotating wiper blades continuously spread the feed along the inner wall of the cylinder to maintain uniformity of thickness and to ensure contact with the heated surface. The volatile components are driven off and collected on an internal chilled condenser surface. The condensate or distillate is removed continuously. At the end of the process, the residual becomes dry and heavy and drops to the bottom of the unit for removal. The wiped-film evaporator is best suited for treatment of viscous or high-solids content feed.

Dryers. Drying, another type of evaporation technique, is suited for waste streams of very high solids content. Several common types of dryers are vacuum rotary dryers, drum dryers, tray and compartment dryers, and pneumatic conveying dryers.

The vacuum rotary dryer is a batch system that uses a cylindrical rotating unit and agitator blades to mix or agitate the waste stream during the drying process. A vacuum is applied and maintained during the process to remove liquids during agitation. The process is terminated when the solids are dried to a specified moisture content.

The drum dryer is a continuously operated unit that uses rotating heated drums for the evaporation contact area. This method of drying can be used with materials high in volatiles. In several designs, the volatiles can be withdrawn and collected for reuse, as appropriate.

A tray and compartment dryer is a batch unit that uses a stationary tray or compartment to dry the waste, generally before transport for disposal or further treatment. Some units can be mounted on removable trucks.

A continuously operating pneumatic conveying dryer is used for applications similar to the tray and compartment dryer. In this case, however, drying is performed in conjunction with grinding, as the solids are conveyed and dried within the unit.

Aqueous Evaporation. Aqueous evaporation for hazardous waste treatment can be accomplished in a closed process vessel that uses steam to evaporate the liquid into a water vapor, which is ultimately condensed and may be reused, as shown in Figure 5. The concentrated liquid is collected for further treatment or disposal.

Ion Exchange. Ion exchange is an adsorption process where ionic species are adsorbed from solution by exchanging places with a similarly charged ion on the exchange media. Ion exchange is used primarily to remove metals, although nonmetallic inorganic and organic ions can also be removed. Metals that can be removed by ion exchange include barium [7440-39-3], cadmium [7440-43-9], hexavalent chromium [7440-47-3], copper [7440-50-8], lead [7439-92-1], mercury [7439-97-6], nickel [7440-02-0], selenium [7782-49-2], silver [7440-22-4], uranium [7440-61-1], and zinc [7440-66-6]. Nonmetallic ions that may be removed by ion exchange include nitrate, sulfate, cyanide (when complexed with iron), and some organic acids, phenols, and amines under special conditions.

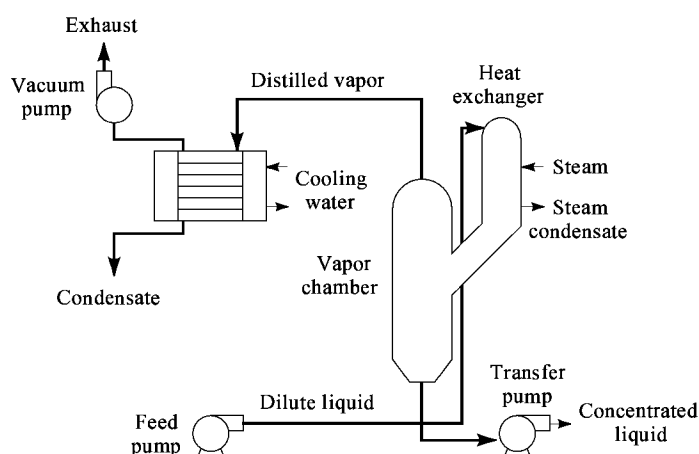


Fig. 5. Example of an aqueous evaporation unit.

The adsorptive materials used for ion exchange are called zeolites. Naturally occurring zeolites belong to a group of hydrous aluminum silicate minerals. Synthetic zeolites or resins are based on organic polymers and are in more common usage today. Ion-exchange resins that adsorb positively charged ions are cationic and those that adsorb negatively charged ions are anionic. These resins can be designed to remove certain ions with greater specificity than other ions; however, ions with the same charge (positive, negative) can still compete for exchange sites and adversely affect the removal efficiency of the targeted ion if they are present in high enough concentrations.

Ion-exchange units can be operated in parallel or series and in downflow or upflow (fluidized) mode. A unit is regenerated when the target ion appears in the effluent (breakthrough). Ion-exchange resins are regenerated by washing with a concentrated solution containing the original exchangeable ion. This results in the desorption of the target ion from the resin with replacement by the original ion. Hydrogen-based cationic resins are regenerated with acid solutions such as sulfuric [7664-93-9], nitric [7697-37-2], or hydrochloric [7647-01-0] acid. Anionic resins that are hydroxide-based are regenerated with caustic solutions such as sodium hydroxide [1310-73-2]. Sodium chloride [7647-14-5] solution is used to regenerate sodium-based cationic and chloride-based anionic resins. Concurrent regeneration is when the regenerant solution flows through the unit in the same direction as does the wastewater, countercurrent regeneration flows in the opposite direction and requires less solution.

Pretreatment of aqueous streams may be required prior to using ion exchange. Suspended solids that can plug an ion-exchange unit should be reduced to the 10 μm level. Organics that can foul resins can be removed by carbon adsorption. Iron [7439-89-6] and manganese [7439-96-5], commonly present in ground waters, should be removed because they precipitate on the resin.

Membrane Filtration. Membrane filtration describes a number of well-known processes including reverse osmosis, ultrafiltration, nanofiltration, microfiltration, and electrodialysis. The basic principle behind this technology is the use of a driving force (electricity or pressure) to filter

particles, ions, and organic molecules through a membrane, producing a clean stream on one side and a concentrated stream on the other. Although membrane filtration is discussed in this section in relation to waste treatment, this technology also has many industrial applications such as water desalination and softening, product concentration and purification, and metal recovery. Applications in waste treatment include metals removal–recovery, oil–water separation, and removal of toxic organic compounds.

The individual membrane filtration processes are defined chiefly by pore size although there is some overlap. The smallest membrane pore size is used in reverse osmosis (0.0005–0.002 microns), followed by nanofiltration (0.001–0.01 microns), ultrafiltration (0.002–0.1 microns), and microfiltration (0.1–1.0 microns). Electrodialysis uses electric current to transport ionic species across a membrane. Micro- and ultrafiltration rely on pore size for material separation, reverse osmosis on pore size and diffusion, and electrodialysis on diffusion. Separation efficiency does not reach 100% for any of these membrane processes. For example, when used to desalinate–soften water for industrial processes, the concentrated salt stream (reject) from reverse osmosis can be 20% of the total flow. These concentrated, yet still dilute streams, may require additional treatment or special disposal methods.

Reverse osmosis and electrodialysis are used to separate metals from electroplating bath rinse waters in the metal finishing industry. Ultrafiltration is used for oils–solids–water separation, metals recovery, and paint recovery from anodic paint processes. It has also been tested as an innovative technology for removal of solids and residual organics following biological treatment of contaminated ground water (4). Microfiltration is used for oil–solids–water separation and metals removal–recovery. Although nanofiltration has been used successfully for water treatment (removal of hardness, organic compounds, and viruses), it has not been used much for waste treatment. Testing has also been conducted with another type of membrane filtration process, pervaporation, that uses a vacuum to induce transfer of volatile organics from a liquid across the membrane (5). Membrane filtration of gaseous streams, which is used on industrial process streams to recover volatile organics, has been tested as posttreatment for air stripping and soil venting gases.

Membranes are subject to fouling which can be caused by metal oxides, precipitating salts, colloids, and biological growth. Cleaning agents include acids, alkalines, oxidizers, detergents, and organic solvents. Pretreatment prior to membrane filtration is required to reduce heavy solids and remove free oil.

Neutralization. Neutralization is pH adjustment of an acidic or caustic waste to a more neutral range. Neutralization is a very common treatment step for wastewaters and gases; it is used less frequently with solid wastes because the pH is less of a problem with these wastes. The most commonly used neutralizing agents are lime or calcium hydroxide [1305-62-0] for acidic wastewaters and sulfuric acid for caustic wastewaters. Other chemicals used for wastewater neutralization include sodium hydroxide and magnesium hydroxide [1309-42-8] for acidic streams and hydrochloric acid and carbon dioxide [124-38-9] for caustic streams. At facilities where both acidic and caustic streams are generated, commingling these streams helps reduce the amount of neutralizing chemicals that would otherwise be required. These acid–caustic waste streams can only be commingled, however, where they are compatible, that is, where the mixture will not generate unwanted precipitants or chemical reactions. Neutralization of acid gases with liquid caustic solutions (wet scrubbing) is very common. For example, the flue gas from an incinerator may be scrubbed with a soda ash [497-19-8] solution to remove acid gases. Wet scrubbing is normally done in a packed tower; the packing referred to is usually small plastic forms that provide a high degree of surface area per unit volume in order to maximize the contact of the gas with the scrubbing solution. There are also dry scrubbing systems for acid gases where the neutralizing agent, for example, lime, is sprayed into the gas stream.

Oil–Water Separation. Gravity separation of oil and water is a very simple process, often used as a pretreatment step in an overall wastewater treatment system. Under quiescent conditions, free oil floats to the surface where it can be skimmed off. Water emulsions may be broken for separation by using coagulants, acids, pH adjustment, heat, centrifuging, or high-potential alternating current. The API separator and corrugated plate interceptor (CPI) are common types of oil–water separators. In the United States, API separator sludge from petroleum refining wastewaters is listed as a hazardous waste, K051. As in the DAF process discussed earlier in this section, it is not the API separator itself that makes K051 hazardous, but the hazardous materials it contains.

Oxidation and Reduction. Oxidation and reduction (redox) reactions are used for both partial and complete degradation of many organic and inorganic compounds. A substance is oxidized when its oxidation state is increased, likewise it is reduced when its oxidation state is reduced. For inorganic compounds, oxidation involves the loss of electrons, reduction is a gain in electrons. Redox reactions for organics are more complicated and may include the transfer of electrons, a hydrogen atom, an oxygen atom, a hydroxyl radical, a chlorine atom, a chlorinium ion (Cl^+), or similar species (6). Examples of compounds that are treated by redox processes include alcohols, phenols, and cyanide. Common chemical oxidants for waste treatment include chlorine [7782-50-5], ozone [10028-15-6], and hydrogen peroxide [7722-84-1].

Oxidizers do not discriminate among compounds and are capable of reacting with any oxidizable compounds in a waste stream. Oxidation is used either to degrade a compound completely or quite often, to degrade a compound partially to a less toxic form or intermediate that can be discharged or if needed, treated further by another process.

Factors affecting oxidation processes include pH, the type and quantity of oxidizable compounds in the waste, and metals that can react with the oxidizing agent. The pH affects the oxidation rate by changing the free energy of the overall reaction, changing the reactivity of the reactants, and or affecting specific OH^- ion or H_2O^+ ion catalysis (6). The presence of oxidizable compounds in addition to the target compound increases the amount of oxidant required. Metal salts, especially those of lead and silver, can also react with the oxidizer and increase the required dosage or interfere with the treatment process.

Chlorine. Chlorine is a well known disinfectant for water and wastewater treatment, however, it can react with organics to form toxic chlorinated compounds such as the trihalomethanes bromodichloromethane, dibromochloromethane, chloroform [67-66-3], and bromoform [75-25-2]. Chlorine dioxide [10049-04-4] may be used instead since it does not produce the troublesome chlorinated by-products as does chlorine. In addition, by-products formed by chlorine dioxide oxidation tend to be more readily biodegradable than those of chlorine, however, chlorine dioxide is not suitable for waste streams containing cyanide.

Cyanide destruction by alkaline chlorination is a widely used process. With alkaline chlorination, cyanide is first converted to cyanate with hypochlorite [7681-52-9] at a pH greater than 10. A high pH is required to prevent the formation of cyanogen chloride [506-77-4] which is toxic and may evolve in gaseous form at a lower pH. With additional hypochlorite, cyanate is then oxidized to bicarbonate, nitrogen gas, and chloride. The pH for this second stage is 7–9.5 (6).

Ozone. A primary advantage of oxidation with ozone is the avoidance of chlorinated by-products of the reaction. Excess ozone that is not consumed in the reaction decomposes to oxygen [7782-44-7]. The main disadvantage is the high electrical cost of producing ozone which is generated onsite from dry air or oxygen by high voltage electric discharge.

Ozone can be used to completely oxidize low concentrations of organics in aqueous streams or partially degrade compounds that are refractory or difficult to treat by other methods. Compounds that can be treated with ozone include alkanes, alcohols, ketones, aldehydes, phenols, benzene and its derivatives, and cyanide. Ozone readily oxidizes cyanide to cyanate, however, further oxidation of the cyanate by ozone proceeds rather slowly and may require other oxidation treatment like alkaline chlorination to complete the degradation process.

Ozone is only slightly soluble in water. Thus, factors that affect the mass transfer between the gas and liquid phases are important and include temperature, pressure, contact time, contact surface area (bubble size), and pH.

Ozonation can be enhanced by the addition of ultraviolet (uv) radiation. This combination can be effective in degrading chlorinated organic compounds and pesticides. In addition, metal ions such as iron, nickel, chromium, and titanium [7440-32-6], can act as catalysts, as can ultrasonic mixing.

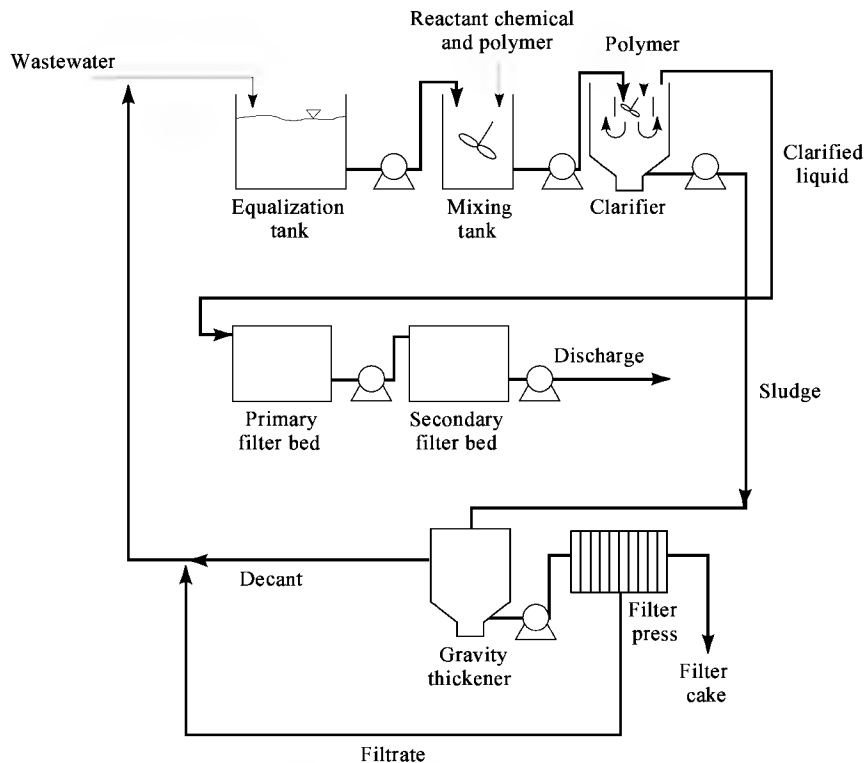


Fig. 6. Example of a chemical precipitation system.

Hydrogen Peroxide. Hydrogen peroxide is typically used as an oxidizer in combination with uv light, ozone, and or metal catalysts. Fenton's reagent is hydrogen peroxide with iron as a catalyst. Hydrogen peroxide uv light has been shown to be effective in oxidizing benzene, chlorobenzene [108-90-7], chloroform, chlorophenol, 1,1-dichloroethane [75-34-3], dichloroethene [540-59-0, 75-35-4], phenol [108-95-2], tetrachloroethylene, 1,1,1-trichloroethane [71-55-6], trichloroethylene, toluene, xylenes (7-10) and many other organic compounds.

Precipitation. Precipitation processes have been used for many years to remove metals from aqueous streams. An example of this process is shown in Figure 6. Metals precipitation is accomplished by pH adjustment and the addition of a chemical reagent that forms a precipitant with the metal that can be settled out and separated from the aqueous stream. Chemical reagents such as calcium hydroxide (lime), sodium hydroxide (caustic), magnesium oxide [1309-48-4], and magnesium hydroxide can be used to precipitate arsenic [7440-38-2], cadmium, trivalent chromium, copper, iron, lead, manganese, nickel, and zinc. Hydroxide precipitation, particularly with lime, is the most common precipitation process because the chemical reagents are readily available, relatively inexpensive, easy to handle, and form sludges that dewater easily. Sulfide precipitation with reagents such as sodium sulfide [1313-82-2] and ferrous sulfide [1317-37-9] can be used with cadmium, cobalt [7440-48-4], copper, iron, mercury, manganese, nickel, silver, tin, and zinc. Carbonate precipitation with reagents such as sodium carbonate (soda ash) and calcium carbonate [471-34-1] can be used with cadmium, lead, nickel, and zinc. Because the sulfide forms of metals are much less soluble than their hydroxide counterparts, sulfide precipitation can treat metals to lower concentrations. Although very insoluble, metal sulfides form very fine, hard to settle particles and generate large quantities of sludge.

The typical precipitation process takes place in a series of tanks beginning with chemical addition, followed by flocculation and coagulation of precipitated solids, sludge thickening, and finally, sludge dewatering. Sand filtration of the wastewater effluent to remove residual solids may be used as a final polishing step. Chemicals may be added to aid flocculation and coagulation and to produce a more dense and easily dewatered sludge. Typical chemical aids are alum (aluminum sulfate [10043-01-3], ferric chloride [7705-08-0], and proprietary polyelectrolytes. Precipitation of arsenic first requires oxidation of the arsenite form to arsenate. Hexavalent chromium requires reduction to the trivalent form.

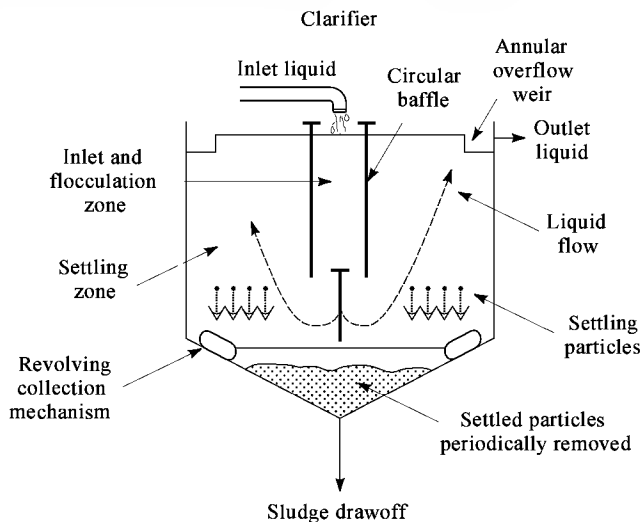


Fig. 7. Example of a sedimentation–clarification system.

Precipitation is affected by pH, solubility product of the precipitant, ionic strength and temperature of the aqueous stream, and the presence of metal complexes. For each metal precipitant, there is an optimum pH where its solubility is lowest and hence, the highest removals may be achieved. When an aqueous stream contains various metals, the precipitation process cannot be optimized for each metal, sometimes making it difficult to achieve effluent targets for each. Solubility products depend on the form of the metal compound and are lowest for metal sulfides, reflecting the relative insolubility of these compounds. For example, the solubility product for lead sulfide [1314-87-0] is on the order of 10^{-45} compared to 10^{-13} for lead carbonate. Metal sulfides and hydroxides are more soluble at higher temperatures, whereas carbonates are less soluble. When metals are complexed or chelated with other ions or molecules (for example, cyanide and ethylenediaminetetraacetic acid (EDTA) [60-00-4], they are more soluble, may not precipitate easily, and require

additional or alternative forms of treatment.

Sedimentation–Clarification. Sedimentation–clarification is a process by which hazardous and nonhazardous grits, fines, and other suspended solids are removed from the waste stream through gravity settling. The process typically is used as a pretreatment step or in conjunction with another treatment process such as chemical or biological treatment. Usually, sedimentation–clarification is performed in a gravity settling clarifier, as shown in Figure 7. A settling system provides enough residence time to allow the heavier suspended solids to gravitationally settle to the bottom of the system. These solids are removed and thickened periodically prior to disposal.

Flocculation, commonly used in clarifiers to enhance sedimentation, is a physical–chemical process in which small particles agglomerate to form larger particles. The large particles, because they are heavier than the smaller particles, settle more effectively in the clarifier or sedimentation basin. Flocculation of certain types of particles can be induced by slow agitation of the wastewater, without the addition of a flocculating agent. However, a flocculating chemical usually is added to the wastewater to promote flocculation of the smaller particles. The flocculants adhere readily to suspended solids and to each other to form larger particles with greater density, which settle better. The settled solids thicken at the bottom of the clarifier and, in some clarifier designs, are used to form a sludge blanket that becomes a filtration mechanism within the clarifier.

Clarifiers typically are used in chemical precipitation and biological treatment processes to remove precipitated metal solids and suspended biological solids. To prevent the sludge blanket from becoming too thick or heavy, part of the sludge blanket is removed continuously or intermittently from the system and thickened prior to disposal.

Solvent Extraction. With solvent extraction, organics are separated from a waste by mixing the waste with a solvent capable of dissolving or extracting the organics. The extraction process can have multiple stages and be operated where the solvent and waste pass concurrently or countercurrently.

After mixing, the solvent and waste are separated. The solvent with dissolved organics is called the extract. The waste remaining after extraction is called the raffinate. The extract may be sent to a distillation or steam stripping unit to separate the dissolved organics from the solvent and the solvent can be recycled back to the extraction process. The raffinate may require additional treatment or may be disposed or incinerated.

Factors affecting solvent extraction include type of solvent and the relative solubilities of the organic components in the solvent, selectivity value (the ratio of the equilibrium concentrations of the organic in the solvent and in the waste), temperature, pH, mixing, and settling time for separation. Temperature and pH affect equilibrium conditions. The type and quantity of mixing affects the degree of contact and transfer efficiency between solvent and waste. Sufficient time must be allowed for settling to separate the extract and raffinate.

Stabilization–Solidification. Stabilization and solidification are technologies that are often used together to improve or strengthen the physical nature of a waste and to bind, immobilize, or otherwise prevent the migration of toxic constituents contained in the waste. Stabilization generally refers to processes that reduce the toxicity, leaching, or mobility of toxic constituents, perhaps through chemical reactions, but the physical form of the waste is not necessarily changed or improved. Solidification generally refers to processes that change the physical nature of the waste to make it more manageable, increase its structural strength and load-bearing capacity, or reduce its permeability, but not necessarily by chemical reaction. Many treatment technologies involve characteristics of both stabilization and solidification, making distinctions between the two processes unclear at times such that the two terms are often used together. Applying the correct definition to a particular treatment process is not important, rather it is in the changes in the waste characteristics that are important. Are toxic constituents rendered less toxic or mobile and can the waste be handled easily and withstand the physical stresses during and after disposal?

Some of the most common stabilization–solidification processes are those using cement, lime, and pozzolanic materials. These materials are popular because they are very effective, plentiful, and relatively inexpensive. Other stabilization–solidification technologies include thermoplastics, thermosetting reactive polymers, polymerization, and vitrification. Vitrification is discussed in the thermal treatment section of this article and the other stabilization–solidification processes are discussed below.

Cement, Lime, and Pozzolanic Materials. Cement is the chemical binder used in making concrete, a chemically cured mixture of cement, water, and aggregate (usually sand and gravel). With sufficient water for hydration, cement forms a calcium aluminosilicate crystalline structure that can physically and/or chemically bind with toxic constituents such as metal hydroxides and carbonates. Use of lime (calcium hydroxide) involves similar reactions where aluminosilicates are supplied by the waste or treatment additives. Pozzolans are siliceous materials that can combine with lime to form cementitious materials (Pozzolans derive their name from Pozzuoli, Italy, near Mt. Vesuvius where a siliceous ash is found). Common pozzolanic materials are fly ash, blast furnace slags, and cement kiln dust. The high pH of cementitious binders protects against acid conditions that can result in metal leaching. Organics interfere with hydration and the formation of the crystalline structure binding the waste. This interference can be minimized by treatment additives such as clays (natural and organically modified), vermiculite, or organically modified silicates. Other waste constituents that can reduce the effectiveness of cementitious processes include sulfates, chlorides, borates, and silt.

Thermoplastics. Wastes containing heavy metals and/or radioactive materials may be mixed with thermoplastics. In this process, the waste is mixed with a molten thermoplastic material such as bitumen, asphalt, polyethylene [9002-88-4], or polypropylene [9003-07-0]. Wastes are normally dried, heated, mixed with the thermoplastic, cooled, and finally placed in containers for disposal. The process is not suitable for organic wastes unless air emission controls are installed because the high temperatures will drive off volatile components. It is also not suitable for hygroscopic wastes which absorb water, expanding and cracking the thermoplastic. Likewise, strongly oxidizing constituents, anhydrous inorganic salts, and aluminum salts are unsuitable for thermoplastics.

Thermosetting Reactive Polymers. Materials used as thermosetting polymers include reactive monomers such as urea–formaldehyde, phenolics, polyesters, epoxides, and vinyls, which form a polymerized material when mixed with a catalyst. The treated waste forms a sponge-like material which traps the solid particles, but not the liquid fraction; the waste must usually be dried and placed in containers for disposal. Because the urea–formaldehyde catalysts are strongly acidic, urea-based materials are generally not suitable for metals that can leach in the untrapped liquid fractions. Thermosetting processes have greater utility for radioactive materials and acid wastes.

Polymerization. Spills of chemicals that are monomers or low-order polymers can be polymerized by adding a catalyst. Compounds that may be treated by polymerization include aromatics, aliphatics, and other oxygenated monomers such as vinyl chloride and acrylonitrile [107-13-1].

Steam Stripping. Steam stripping is used to separate volatile components from wastewater. Steam stripping is very similar to distillation which has been discussed earlier. Steam stripping differs from distillation, however, in that there is no rectifying section and the waste steam is introduced at the top of the column (packed or tray). Steam is introduced at the bottom and as it rises countercurrent to the downflowing wastewater, it vaporizes the volatile components. The vapor stream is collected overhead and condensed. Steam stripping is amenable to volatile components that phase separate from water. The organic and aqueous phases are separated after condensation; the organics may be recovered or disposed and the aqueous stream can be recycled to the stripper. A column with rectification (ie, distillation) is required if the volatile components are not phase-separable in order to obtain a high concentration of volatiles in the overhead stream.

Supercritical Fluid Extraction. Supercritical fluid extraction takes advantage of the enhanced solvent power of compounds at supercritical temperature and pressure conditions to extract organics from oily and/or organic liquids and sludges. Carbon dioxide [124-38-9] is a particularly attractive solvent at supercritical conditions (greater than 7400 kPa and 31°C) because it is nontoxic and leaves no residue when it is reverted to gaseous form during the recovery step. Carbon dioxide is not suitable for wastes with high alkalinity, however, because excessive amounts of the supercritical solvent will be consumed in the formation of carbonates and bicarbonates. Applications using carbon dioxide have been developed to clean products and equipment to replace toxic organic solvents.

Supercritical Water Oxidation. Supercritical water oxidation is similar to wet air oxidation in that it uses high temperature and high pressure conditions to oxidize organics. Temperature and pressure conditions, however, are higher (greater than 22,000 kPa and 374°C) such that water is at supercritical conditions where it has properties between those of a liquid and a gas. Organics are easily solubilized in supercritical water and once solubilized, they are completely oxidized at the high temperature and pressure conditions. Reaction times are very short (a few seconds to less than a minute) and treatment efficiencies very high (essentially 100%). Supercritical water oxidation has the capability of treating many different kinds of organics including polycyclic aromatic hydrocarbons, chlorinated hydrocarbons, PCBs, paint, oil, dyes, pulp and paper wastes, chemical warfare agents, and missile propellants (11). Wastes must be either aqueous or in slurry form.

Reaction vessels for supercritical water oxidation must be highly corrosion resistant because of the aggressive nature of supercritical water and oxidation reaction products at extreme temperatures and pressures. Supercritical oxidation of PCBs and some chlorinated hydrocarbons can be difficult

because acids are formed that are highly corrosive to almost any materials used for reactor components (11). Inorganic salts can be a problem because they are sparingly soluble under these conditions and will precipitate on the vessel walls and components.

Wet Air Oxidation. With wet air oxidation, increased temperature and pressure are used to oxidize dilute concentrations of organics and some inorganics, such as cyanide, in aqueous wastes that contain too much water to be incinerated, but are too toxic to be treated biologically. In general, wet air oxidation provides primary treatment for wastewaters that are subsequently treated by conventional methods. This technology can be used with wastes that are pumpable (slurries and liquids).

Waste streams that are treated by wet air oxidation generally are those having dissolved or suspended organic concentrations from 500 to 50,000 mg L. Below 500 mg L, oxidation rates are too slow and above 50,000 mg L, incineration may be more feasible.

Operating parameters include temperature, pressure, oxygen concentration, and residence time. Materials of construction include stainless steel, nickel, and titanium alloys (the latter for extremely corrosive wastes containing heavy metals). Vented gases from the process may require scrubbing or other emission controls.

Because of high energy input, wet air oxidation is relatively expensive. To reduce energy demand, the incoming wastewater can be preheated by heat exchange with the treated effluent. Operating temperatures may be lowered by using a catalyst. Typically, removal of the catalyst from the effluent is needed to avoid contamination and to minimize operating costs.

Biological Treatment

Biodegradability of Compounds. Biological processes are effective in treating a wide variety of organic and inorganic compounds. Most biological processes are aerobic and use the oxygen available in supplied air or pure oxygen. The chief products of aerobic biodegradation of carbonaceous and nitrogenous compounds are carbon dioxide, ammonia, water, and biomass (cell growth of the microorganisms). The ease with which a compound is degraded is related to its chemical structure. Some compounds are not degraded completely, but are instead broken down into simpler compounds.

The biodegradability of hazardous compounds according to general structure has been summarized in Ref. 12 from a number of studies. They have found that the biodegradability of hydrocarbons generally decreases in the following order: alkanes and alkenes with straight alkyl chains, mono- and dicycloalkanes, and aromatic hydrocarbons, including polycyclic ones. Mono- and dicyclic aromatic hydrocarbons are usually more biodegradable than cycloalkanes. Chain branching in alkanes and alkenes decreases biodegradability, although biodegradability of alkanes and alkenes are essentially the same if they have identical chain configurations. Monocycloalkanes are usually readily biodegradable, although exceptions have been noted with specific compounds. Polycycloalkanes are less biodegradable, especially tetra- and higher polycycloalkanes. Naphthalene [91-20-3], lower alkylbenzenes, and phenanthrene [85-01-8] are aromatic hydrocarbons that are biodegraded relatively easily. The biodegradability of alkylbenzenes depends on the number and position of the alkyl groups. Polyalkylation of aromatic hydrocarbons usually decreases biodegradability. With phenylalkanes, biodegradability depends on the branching of long alkyl chains and those phenylalkanes with a quaternary carbon atom, particularly in the terminal position, are quite resistant to biodegradation. Polycyclic aromatic hydrocarbons (except for naphthalene) are relatively slow to biodegrade, particularly tetra- and higher aromatics.

General guidelines for biodegradability have been developed by others. Ref. 13 gives the following general guidelines for the biological treatment of industrial wastewaters:

Nontoxic aliphatic compounds containing carboxyl, ester, or hydroxyl groups are readily biodegradable. Those with dicarboxylic groups require longer acclimation times than those with a single carboxyl group.

Compounds with carbonyl groups or double bonds are moderately degradable and slow to acclimate.

Compounds with amino or hydroxyl groups decrease in biodegradability relative to degree of saturation, in the following order: primary, secondary, then tertiary carbon atom of attachment.

Biodegradability decreases with increasing degree of halogenation.

Ref. 14 contains a more general and simplistic order of decreasing biodegradability: straight-chain compounds, aromatic compounds, chlorinated straight-chain compounds, and chlorinated aromatic compounds.

Many different factors influence the performance of biological treatment systems. Although each specific biological process has special requirements, factors common to biological processes, besides biodegradability, include: organic concentration, temperature, pH, nutrients, and oxygen (aerobic or anaerobic).

Activated Sludge. Activated sludge is widely used to treat both municipal and industrial wastewater. Used in a variety of process configurations and operating modes, the activated sludge process contains three basic elements: an aeration basin, clarification, and sludge recycle. The aeration basin contains a suspension of *activated* microorganisms, primarily bacteria and protozoa, that biodegrade the wastewater under aerobic conditions. Aerators supply oxygen for the microorganisms and mixing energy to keep them in suspension and in contact with the waste. The microorganisms are separated from the treated wastewater by clarification and a portion of this biological sludge is recycled back to the aeration basin to maintain the microbial population and begin a new treatment cycle.

Although the activated sludge process is relatively robust and able to handle variable wasteloads and operating conditions, it must be protected from shock loadings that are toxic or excessively high in concentration. Influent variability, both in flow and quality, can be reduced through equalization prior to the aeration basin. Other pretreatment steps that may be needed to protect the aeration basin are neutralization, oil-water separation, and grit or solids removal. An example of an activated sludge process is shown in Figure 8.

The microorganisms grow in response to the food source supplied in the wastewater and produce more biological sludge than is needed to maintain the process. This excess sludge must be wasted from the process and is usually treated by dewatering and aerobic or anaerobic digestion.

Aeration basins can be constructed as concrete or steel tanks or earthen impoundments, although tanks are more common in the United States now because of ground water problems with leakage from impoundments and stringent regulation of impoundments for the treatment of hazardous waste.

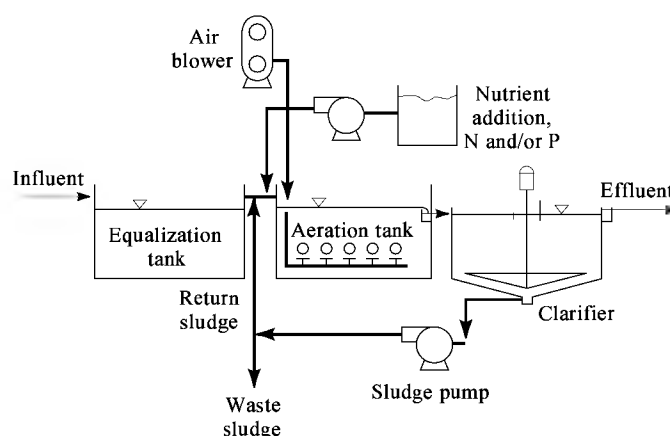


Fig. 8. Example of an activated sludge process.

Process Variations. There are many variations of the activated sludge process which have been developed in response to different wastewater

characteristics and process requirements. The mixing regime of the process may be plug flow, completely mixed, or more often, something between the two. In a plug flow system, influent wastewater and return activated sludge are mixed at the head of the aeration basin and travel as a *plug* along the length of the basin. Therefore, the organic concentration is greatest at the head of the basin and diminishes as the wastewater travels towards the basin outlet. In a completely mixed activated sludge system, the influent wastewater and return activated sludge are uniformly or completely mixed in the aeration basin. Mixing the influent wastewater with the larger volume of the aeration basin allows high organic loads and toxic or inhibitory compounds to be diluted to protect the system from shock, an advantage over plug flow systems. In general, complete mix systems have higher mixed liquor concentrations than plug flow systems and consequently can treat high organic loads. However, complete mix systems tend to produce a sludge that is less dense and may be more difficult to settle. In reality, pure plug flow and complete mixing do not exist as the process operates somewhere in between. Some of the most common process variations of activated sludge systems are described in the following sections.

Tapered Aeration. In this process, aeration is provided in a tapered or stepped fashion to match oxygen requirements in a plug flow system. Oxygen requirements are highest at the head of the aeration basin where the organic-laden influent enters. Aeration is decreased along the basin as the organics are degraded. Toward the end of the basin, it may be necessary to maintain a certain level of aeration in excess of oxygen requirements in order to provide enough mixing energy to keep the microorganisms in suspension.

Contact Stabilization. This process takes advantage of the ability of the activated sludge to quickly adsorb organics. Contact stabilization is operated either as a two-unit system or as two zones within a single aeration basin. In the two-unit system, the influent wastewater is first mixed with return activated sludge in a contact basin. Contact time is short, 30–90 min, but sufficient for the sludge to adsorb the organics. The mixture then flows to a clarifier where the sludge and adsorbed material are settled out. The treated effluent is discharged from the clarifier. A portion of the settled sludge is wasted as excess biomass and the remainder enters the stabilization basin where the adsorbed organics are biodegraded over a period of 3–6 h, thereby reactivating the sludge. In the two-zone system, the aeration basin is operated as plug flow. The contact zone for the influent wastewater is near the end of the aeration basin where organic concentrations are lowest and the sludge is ready to adsorb additional organics. The stabilization zone begins at the head of the aeration basin where the settled sludge with its adsorbed organic load enters. Contact stabilization requires about 50% less aeration basin capacity than conventional systems and produces less biomass which settles better. On the down side, effluents are poorly nitrified and the process is better suited to wastewaters that contain high proportions of biochemical oxygen demand (BOD) as colloidal or suspended matter.

Extended Aeration. As its name implies, extended aeration operates with a long aeration period. With a long sludge age and low food-to-microorganism ratio, sludge production is low and the sludge settles well. Because aeration and basin volume requirements are high, extended aeration is normally used for small flows and small communities. The oxidation ditch is a common form of the extended aeration process. Viewed from above, an oxidation ditch looks like a racetrack around which the wastewater flows. Aeration is provided by rotors or brushes which also provide the energy to keep the water flowing around the track. Influent wastewater usually enters upstream of the aerator. Anoxic zones can be set up in oxidation ditches, allowing nitrogen removal through denitrification. In addition to the racetrack design, oxidation ditches can be of the carrousel type which resembles a long racetrack folded in half. The carrousel has two aerators that allow larger flows to be treated.

Pure Oxygen. Activated sludge systems can be designed to use pure oxygen instead of air to increase oxygen transfer rates. Advantages of pure oxygen systems are increased BOD removal rates, smaller aeration basin volumes, lower sludge production, and improved sludge settling. The aeration basin is covered and divided into a series of stages. Influent wastewater, return activated sludge, and oxygen are introduced into the first stage. In each progressive stage, both the BOD and oxygen content decrease, acting like a tapered aeration plug-flow system. Operating parameters for pure oxygen systems, such as mixed liquor suspended solids, mixed liquor oxygen levels, and food-to-microorganism (F/M) loading rate, are higher than for conventional activated sludge. Disadvantages of pure oxygen systems include the high cost of oxygen production and the need for safety precautions and monitoring with combustible hydrocarbons in high oxygen environments.

PACT System. Powdered activated carbon treatment, or PACT (is a registered trademark of Du Pont), is the simple addition of powdered activated carbon to an activated sludge system. This process is used to prevent inhibitory effects of toxic compounds, remove refractory organics or color, or remove effluent toxicity. Other benefits of the process are reduced emissions of volatile compounds from wastewaters and improved sludge settling. Carbon dosage rates depend on the given situation, but usually range from 50 to 3000 mg/L.

Process Control. There are many different process control variables that determine the efficiency of the activated sludge process. Effluent quality depends primarily on wastewater characteristics, plant loading (organics and flow), residence time in the system, temperature, type and concentration of microorganisms degrading the waste, oxygen supply, and degree of mixing. The main process control variables are described below.

Hydraulic Retention Time. The hydraulic retention time (HRT) is a measure of the average time the wastewater spends in the aeration basin. It is calculated by dividing the volume of the aeration basin by the influent wastewater flow rate. Since the flow rate of recycle sludge is not included in the calculation, the actual retention time is less. The HRT depends on the type of activated sludge process being used. Conventional systems have HRTs at 4–8 h. Extended aeration systems can have much longer HRTs: 18 h to as much as 72 h.

Organic Loading. Organic loading is the mass of influent BOD per volume of aeration basin. Typical values for conventional systems are 0.32–0.64 kg/1000 L. Loading for extended aeration systems are lower whereas loadings for most other process configurations are higher.

MLSS/MLVSS. The concentration of microorganisms in the aeration basin determines the organic loading and the efficiency of the process. The contents of the aeration basin, the mixture of wastewater and microorganisms, is referred to as mixed liquor. The mixed liquor suspended solids (MLSS) and mixed liquor volatile suspended solids (MLVSS) are both used as measures of the microorganism concentration in the aeration basin. The MLVSS more closely approximates the biologically active portion of the solids in the mixed liquor because microbial cellular material is organic and volatilizes or burns at 500°C (the temperature for volatile solids analysis). The MLSS includes both the volatile and inert solids in the mixed liquor. It is often used or at least monitored more frequently than the MLVSS because the analytical test for suspended solids is quicker than for volatile suspended solids. The volatile fraction representing the MLVSS varies, but a typical value is 0.8. The MLSS concentration for conventional systems ranges from 1500 to 3500 mg/L. MLSS concentrations will be higher in other process configurations. For example, pure oxygen systems may have MLSS concentrations from 5000 to 8000 mg/L. Specially acclimated systems may have MLSS concentrations of 10,000 mg/L and higher. MLSS concentrations are limited by the availability of oxygen in the aeration basin, and recycle sludge flow rates.

F/M Ratio. The food-to-microorganism ratio (F/M) is a measure of the organic loading on the biological population. The F/M ratio is calculated as the mass of BOD per day divided by the mass of MLSS or MLVSS in the aeration basin. Thus, the F/M ratio has units of day⁻¹. The F/M ratio can be controlled by adjusting both the flow rate and concentration of the sludge recycled to the aeration basin. When based on MLVSS, typical values for the F/M ratio for conventional systems range from 0.2 to 0.4 day⁻¹. The F/M ratio is higher for contact stabilization (0.2 to 0.6 day⁻¹) and pure oxygen systems (0.4 to 1.0 day⁻¹), and lower for extended aeration systems (0.03 to 0.15 day⁻¹).

Mean Cell Residence Time. The mean cell residence time (MCRT) is a measure of the average length of time (in days) spent in the system by the microorganisms and relates to the overall growth phase of the microbial population. It is calculated as the mass of microorganisms in the aeration basin, as represented by the MLSS or MLVSS, divided by the mass of microorganisms removed from the system through sludge wastage and in the effluent discharge. The MCRT may also be referred to as the solids retention time or sludge residence time (SRT) or sludge age. The MCRT in conventional systems typically ranges from 3 to 5 days. Pure oxygen systems have higher MCRTs (8 to 20 days) as do extended aeration systems (6 to 40 days).

Sludge Recirculation. The concentration of microorganisms in the aeration basin is achieved by recirculating settled sludge from the clarifier. Both the flow rate and concentration of the return activated sludge determine the solids in the aeration basin. The recirculation rate is the ratio of the flow rate of the return activated sludge to the flow rate of the influent wastewater. Typical ratios for conventional systems are 0.25 to 0.5. Pure oxygen systems have similar rates, whereas the recirculation rates are higher for both contact stabilization (0.25 to 1.0) and extended aeration (0.75 to 1.5).

Aeration. Aeration not only delivers oxygen to the activated sludge system, but provides mixing energy to keep the microorganisms in suspension and in contact with the wastewater. Aeration is provided by mechanical aeration using surface aerators or by diffusers on the bottom of the aeration basin. Both types of aeration equipment are common in conventional and pure oxygen systems, whereas horizontal surface aerator brushes are used in oxidation ditches. In activated sludge systems using air, oxygen concentrations in the aeration basin are usually maintained between 0.5 and 2.0 mg/L. In pure oxygen systems, the concentrations are much higher (4 to 8 mg/L).

Thermal Treatment

Thermal treatment is used to destroy, break down, or aid in the desorption of contaminants in gases, vapors, liquids, sludges, and solids. There are a variety of thermal processes that destroy contaminants, most of which are classified as incineration. Incineration literally means to become ash (from Medieval Latin, *incinerare*, in or into ashes). With respect to the incineration of hazardous wastes regulated in the United States, however, there is a strict legal definition of what constitutes an incinerator. The Resource Conservation and Recovery Act (RCRA) definition of incinerator at 40 *CFR* §260.10 is

any enclosed device that: (1) uses controlled flame combustion and neither meets the criteria for classification as a boiler, sludge dryer, or carbon regeneration unit, nor is listed as an industrial furnace; or (2) meets the definition of infrared incinerator or plasma arc incinerator.

A more simplified description is a unit that combusts materials in the presence of oxygen at temperatures normally ranging from 800 to 1650°C. A typical configuration of an incinerator is shown in Figure 9. Typical types of incineration units that are discussed herein are catalytic oxidation, fluidized beds, liquid injection, multiple hearth furnaces, and rotary kiln. Thermal desorption is also discussed. However, an overview of the main factors affecting incinerator performance is presented first, below.

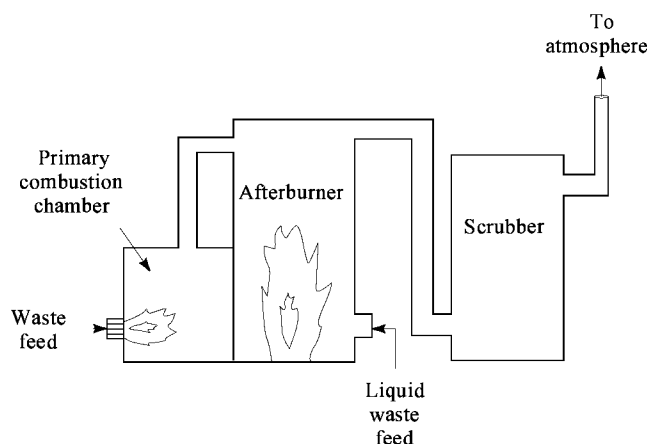


Fig. 9. Typical configuration of an incinerator.

Factors Affecting Performance. There are many factors that affect both the choice of a particular thermal treatment and its performance. Chief among these are waste characteristics, temperature, residence time, mixing or turbulence, and air supply.

Waste Characteristics. The physical form of the waste restricts the applicability of some thermal technologies. A convenient summary of physical characteristics and thermal processes has been presented by Cheremisinoff (15). Those processes that can treat the widest range in physical form are incineration, pyrolysis, calcination, and microwave discharge. These processes are capable of treating wastes in essentially any form: containerized, solid, sludge, slurry, liquid, or fume. Molten salt reactors and plasma technology can be used with any of these types of wastes, except fumes. Evaporative processes are amenable to solids, sludges, slurries, and liquids. Wet air oxidation can be used for slurries and liquids. Distillation (with and without steam) and steam stripping are limited to liquid treatment. Catalytic incineration is limited to fume treatment.

Other waste characteristics that affect the choice of thermal treatment and its operation include heating value, contaminant boiling points, metal content, heterogeneity, moisture content, decomposition products and by-products, and the presence of explosive constituents. The heating value and moisture content of the waste determine if combustion is self-supporting or if auxiliary fuel will be required. Elemental constituents in the waste may be carried in the exhaust gases and must be removed. Problem metals are antimony [7440-36-0], arsenic, cadmium, lead, and mercury. Acid gases containing chlorine, sulfur [7704-34-9], fluorine [7782-41-4], or bromine [7726-95-6] that are generated must be neutralized and scrubbed from the flue gases. The heterogeneity of the waste is determined by the distribution and concentration of combustible organics; localized concentrated materials can produce hot spots in the combustion chamber. Heterogeneity is also defined by the type and amount of debris that may be included in excavated wastes. These materials can interfere with the process by shielding the waste, fouling equipment, or slagging in the reactor. Undesirable decomposition products and byproducts include nitrogen oxides and dioxins.

Temperature. The temperature for combustion processes must be balanced between the minimum temperature required to combust the original contaminants and any intermediate by-products completely and the maximum temperature at which the ash becomes molten. Typical operating temperatures for thermal processes are: incineration (750–1650°C), catalytic incineration (315–550°C), pyrolysis (475–815°C), and wet air oxidation (150–260°C at 10,350 kPa) (15). Pyrolysis is thermal decomposition in the absence of oxygen or with less than the stoichiometric amount of oxygen required. Because exhaust gases from pyrolytic operations are somewhat "dirty" with particulate matter and organics, pyrolysis is not often used for hazardous wastes.

Residence Time. For cost efficiency, residence time in the reactor should be minimized, but long enough to achieve complete combustion. Typical residence times for various thermal processes are incineration (0.1 s to 1.5 h), catalytic incineration (1 s), pyrolysis (12–15 min), and wet air oxidation (10–30 min) (15).

Turbulence. Turbulence is important to achieve efficient mixing of the waste, oxygen, and heat. Effective turbulence is achieved by liquid atomization (in liquid injection incinerators), solids agitation, gas velocity, physical configuration of the reactor interior (baffles, mixing chambers), and cyclonic flow (by design and location of waste and fuel burners).

Air Supply. Oxygen in excess of stoichiometric requirements for complete combustion is needed because incineration processes are not 100% efficient and excess air is needed to absorb a portion of the combustion heat to control the operating temperature. In general, units that have higher degrees of turbulence such as liquid injection incinerators require less excess air (20 to 60%) while units with less mixing such as hearth incinerators require more (30 to 100%).

Catalytic Oxidation. Catalytic oxidation is used only for gaseous streams because combustion reactions take place on the surface of the catalyst which otherwise would be covered by solid material. Common catalysts are palladium [7440-05-3] and platinum [7440-06-4]. Because of the catalytic boost, operating temperatures and residence times are much lower which reduce operating costs. Catalysts in any treatment system are susceptible to poisoning (masking of or interference with the active sites). Catalysts can be poisoned or deactivated by sulfur, bismuth [7440-69-9], phosphorus [7723-14-0], arsenic, antimony, mercury, lead, zinc, tin [7440-31-5], or halogens (notably chlorine); platinum catalysts can tolerate sulfur compounds, but can be poisoned by chlorine.

Fluidized Beds. In this type of incinerator, the waste and an inert bed material, sand normally, or alumina (aluminum oxide) [1344-28-1], are fluidized by blowing heated air through a distributor plate at the bottom of the bed. Fluidization results in good mixing and uniform distribution of materials within the bed which results in lower operating temperatures (450–710°C) and lower excess air requirements. Fluidized beds can be used to incinerate gases, liquids, or solids, although the each type of waste is introduced into the bed differently. Fluidized beds are susceptible to seizing or binding

up or agglomeration on individual bed particles which can lead to seizing up. Bed seizure can occur with wastes containing clays, inorganics (salts), or high concentrations of lime.

Liquid Injection. Liquid injection units are the most common type of incinerator today for the destruction of liquid hazardous wastes such as solvents. Atomizers break the liquid into fine droplets (100–150 microns) which allows the residence time to be extremely short (0.5–2.5 s). The viscosity of the waste is very important; the waste must be both pumpable and capable of being atomized into fine droplets. Both gases and liquids can be incinerated in liquid injection units. Gases include organic streams from process vents and those from other thermal processes; in the latter case, the liquid injection incinerator operates as an afterburner. Aqueous wastes containing less than 75% water can be incinerated in liquid injection units.

Multiple Hearth Furnaces. Although well-suited for combustion of wet sludges, particularly municipal biological wastewater sludges, multiple hearth incinerators are not often used for hazardous wastes because operating temperatures (800–1000°C) of this type of incinerator are too low for many hazardous compounds. The multiple hearths are stacked vertically and sludges are introduced unto the top hearth. A series of "rabble" arms extending from a center rotating shaft mix and move the waste along the circular hearth until the waste falls through holes unto the hearth below. For incineration, usually a minimum of six hearths is required; more hearths are required for pyrolysis. Liquid wastes may be incinerated along with sludges when injected through burner nozzles.

Rotary Kiln. A rotary kiln is a long, cylindrical incinerator that is sloped a few degrees from horizontal. Waste is introduced at the upper end. The gentle slope and slow rotation of the kiln continually mix and re-expose the waste to the hot refractory walls, moving the waste toward the exit point. Rotary kilns operate in continuous or batch mode, although continuous operation is more cost-efficient and less wearing on the refractory material. Operating temperatures range from about 800 to 1650°C. They can be operated in either ashing or slagging mode; the former is common in the United States. Rotary kilns are relatively large incinerators, but because of their versatility in treating a wide range of wastes and their treatment efficiency, they are often used for incinerating hazardous wastes. Rotary kilns can be used to incinerate gases, liquids, sludges, and solids. However, highly aqueous organic sludges tend to form a ring on the walls of the kiln that prevent the discharge of ash and they can also form sludge balls that are not completely incinerated.

Thermal Desorption. Thermal desorption is an innovative treatment that has been applied primarily to soils. Wastes are heated to temperatures of 200 to 600°C to increase the volatilization of organic contaminants. Volatilized organics in the gas stream are removed by a variety of methods including incineration, carbon adsorption, and chemical reduction.

Soil and Ground Water Treatment

Contaminated soil and ground water present special challenges because the contamination is often diffuse and difficult to access. *In situ* processes, those that leave both soil and ground water in place during treatment, are often favored because they are less disruptive; they take advantage of natural, existing conditions; and they preserve the existing subsurface environment. *Ex situ* and pump and treat processes remove soil and ground water for treatment. They may be favored over *in situ* treatment where they will reduce cleanup times, their operation and capabilities are considered more reliable or better understood, or they can achieve lower cleanup levels. Both *in situ* and *ex situ* treatment for soil and ground water rely on a combination of unit processes, which often include biological degradation of organics.

Nonaqueous phase liquids (NAPLs) present special problems for soil and ground water cleanup. Contaminant transport through ground water depends in part on the water solubility of the compound. Because NAPLs cling to subsurface particles and are slow to dissolve in ground water, they hinder cleanups and prolong cleanup times. Dense nonaqueous phase liquids (DNAPLs) migrate downward in the aquifer and can collect in pools or pockets of the substructure. Examples of DNAPLs are the common solvents tetrachloroethylene (PCE) and trichloroethylene (TCE) which were used extensively at many facilities before the extent of subsurface contamination problems was realized.

Today, strict regulation of hazardous waste and material management promises to minimize future soil and ground water contamination, however, past practices have left the environment with a large legacy of contaminated sites that are often difficult and costly to remediate. This is the driving force behind the development of innovative technologies for soil and ground water cleanup. The United States Environmental Protection Agency (U.S. EPA) strongly supports the development of innovative technologies and maintains a Vendor Information System for Innovative Treatment Technologies (VISITT) database to provide information on availability, performance, and cost of these technologies.

In this section, soil and ground water cleanup technologies are divided into five subsections. Many of the unit processes for soil and ground water treatment are general waste treatment technologies, but are discussed here in the context of soil and ground water cleanup. A more general discussion of these technologies are found elsewhere in this article.

Plume Containment. Wells can be placed at a contaminated site to prevent the contamination from spreading further or migrating offsite. In the past, containment efforts often relied on physical methods such as bentonite slurry trenches, grout curtains, sheet piling, well points, and fixative injections. Containment by judiciously placed wells generally costs less, takes less time to install, and is more flexible because pumping rates and locations can be varied.

Four typical well patterns for contaminant plume containment are described in Ref. 16. The first is a pair of injection-production wells. The second is a line of downgradient pumping wells. The third is a pattern of injection-production wells around the boundary of a plume. The fourth, the *double-cell* system, uses an inner cell and outer recirculation cell, with four cells along a line bisecting the plume in the direction of flow. Two other methods of plume containment are biofilters and a funnel-and-gate system, which are described in the *in situ* bioremediation section.

Bioremediation. Bioremediation has great appeal. It is a natural process that degrades hazardous organic chemicals into innocuous carbon dioxide and water or nonhazardous byproducts and it is often less expensive and more effective than pump and treat methods. Articles on bioremediation appear regularly in environmental journals and the U.S. EPA has its own regular series of reports on current activities called "Bioremediation in the Field."

On the down side, there is lot that is not known or completely understood about the process that makes it difficult to design and prove up in field applications. It is a relatively slow process and complete remediation of a site can take several years. The efficiency of the bioremediation cannot always meet target cleanup levels. Bioremediation has been very successful in cleaning up sites contaminated with gasoline and other fuels because the primary contaminants, benzene, toluene, ethylbenzene, and xylenes (BTEX), are relatively easy to extract and degrade. Other compounds have been biodegraded successfully, however, there are many compounds that are still subjects of bioremediation research. Nevertheless, what is known and has been done so far with bioremediation is very encouraging and points to the great potential that bioremediation has in the future of contaminated ground water and soil.

Biodegradation Principles. Microorganisms degrade organic compounds to synthesize cellular materials and to obtain the energy to grow and reproduce. Anabolism results in the synthesis of new cellular material and catabolism (degradation) produces chemicals with lower free energy, and in the case of aerobic biodegradation, often results in complete decomposition of a hydrocarbon to carbon dioxide and water. The principal microorganisms that are involved in the degradation of organic chemicals in water and soil are bacteria and fungi.

Microorganisms can degrade compounds under aerobic or anaerobic conditions, although the types of microorganisms and compounds and degradation process and products may differ. In aerobic environments, free oxygen is present. In anaerobic conditions, free oxygen is not present although oxygen may be present in the form of compounds such as nitrates, sulfates, and iron oxide.

Respiration, or biological oxidation, is the use of oxygen as an electron receptor in the catabolic degradation of an organic and can occur either aerobically or anaerobically. Aerobic respiration uses free oxygen as an electron receptor whereas anaerobic respiration uses inorganic oxygen. In both cases, however, water and carbon dioxide are the principal end products.

Fermentation is an anaerobic catabolic process that uses organics as electron receptors. Since fermentation produces organic products that have lower free energy than their precursors, it is useful in remediation. The lowest free energy form of carbon produced is methane [74-82-8].

Although both fermentation and respiration degrade organics, microorganisms that use respiration to meet their energy requirements for synthesis and reproduction generally grow much more quickly. This, and the fact that respiration completely oxidizes organics to carbon dioxide and water, usually make aerobic biodegradation the treatment of choice. However, because certain synthetic organics do not or are difficult to degrade aerobically, there is much interest in anaerobic processes.

Cometabolism refers to situations where a compound cannot be biodegraded effectively unless another food source is available. The recalcitrant compound, such as TCE, does not provide the energy to allow the microorganisms to grow and thrive. When another food source is available such as

methane, energy is produced for growth and enzymes are produced to metabolize the energy-providing food source and cometabolize the recalcitrant compounds.

Types of Treatment. Bioremediation can be conducted either *ex situ* or *in situ*. *Ex situ* is where ground water is pumped to the surface or soil is excavated for treatment. Pump and treat refers to the *ex situ* removal and treatment of ground water. After treatment, the ground water can either be reinjected or discharged; treated soils can be redeposited, landfilled, or recycled. *In situ* bioremediation takes place in the ground; neither ground water or soil is extracted although ground water may be pumped to the surface for oxygen and nutrient addition. *In situ* bioremediation has become very popular because the process takes full advantage of natural conditions (microorganisms, soil, water) as a bioreactor, little site disturbance is necessary, and it is less equipment intensive than *ex situ* treatment.

Ex situ bioremediation may use various biological wastewater treatment processes, soil piles, or land application. With *in situ* bioremediation, the basic process is the same: microbes, soil, and water working together as a bioreactor. Where the *in situ* techniques differ are in how contaminants and microbes are brought in contact and how oxygen, nutrients, and other chemical supplements are distributed in the soil–water–air matrix. Typical *in situ* bioremediation techniques include natural or intrinsic attenuation, air sparging, and bioventing.

In Situ Bioremediation. *In situ* bioremediation can be an aerobic or anaerobic process, or a combination of the two. In designing an *in situ* bioremediation system, one should consider the types of microorganisms available (naturally in place or added), the structural and chemical makeup of the soil matrix, types of contaminants, oxygen and nutrient addition and distribution, and temperature. These factors are discussed prior to introducing the individual techniques for *in situ* bioremediation.

Aerobic, Anaerobic, and Combined Systems. The vast majority of *in situ* bioremediations are conducted under aerobic conditions because most organics can be degraded aerobically and more rapidly than under anaerobic conditions. Some synthetic chemicals are highly resistant to aerobic biodegradation, such as highly oxidized, chlorinated hydrocarbons and polynuclear aromatic hydrocarbons (PAHs). Examples of such compounds are tetrachloroethylene, TCE, benzo(a)pyrene [50-32-8], PCBs, and pesticides.

Besides being slower, anaerobic treatment is more difficult to manage and can generate by-products that are more mobile or toxic than the original compound, for example, the daughter products of TCE, ie, dichloroethenes and vinyl chloride. It requires a longer acclimation period which means slower startup times in the field. The microbial processes are less well understood, and hence, are less controlled than for aerobic systems.

Nevertheless, an anaerobic system may be the method of choice under certain conditions: (1) contamination with compounds that degrade only or better under anaerobic conditions, (2) low yield aquifers that make pump and treat methods or oxygen and nutrient distribution impractical, (3) mixed waste contamination where oxidizable compounds drive reductive dehalogenation of chlorinated compounds, or (4) deep aquifers that make oxygen and nutrient distribution more difficult and costly.

Anaerobic respiration can degrade organics by using nitrate or sulfate as oxygen sources. Compounds that have been shown to be biodegraded by nitrate-respiring microorganisms are methanes, carbon tetrachloride [56-23-5], *m*-xylene, and some phenols, cresols, and PAHs. Chloroform and stable chlorinated ethanes and ethenes are not degraded by nitrate respiration. Reductive dehalogenation, not degradation, has been observed of naturally occurring aromatics where sulfate respiration was occurring. Sulfate respiration can be inhibited by the accumulation of its byproduct, hydrogen sulfide [7783-06-4].

Combined aerobic–anaerobic systems use sequenced aerobic and anaerobic conditions to degrade compounds that are resistant to aerobic biodegradation. Ground water passes through the anaerobic zone first which partially degrades the resistant compounds. These degradation products are then further degraded in an aerobic zone. A two-zone, plume interception approach was tested under the U.S. EPA's Superfund Innovative Technology Evaluation (SITE) Emerging Technologies Program to degrade chlorinated compounds (17). The first zone was anaerobic where methanogenic (methane producing) microorganisms were expected to promote reductive dechlorination of chlorinated solvents. The second zone was aerobic where the partially dechlorinated products were to be biodegraded. Combined systems do not necessarily have to be separate zones in the aquifer. Aerobic and anaerobic conditions can be alternated by eliminating the oxygen source at timed intervals.

In one study, a coarse-grained sand aquifer was injected with methane and oxygen to stimulate the production of methane monooxygenase (MMO) enzyme which is capable of degrading TCE (18). TCE, added at 60–100 µg L, was degraded by 20–30%. Injected concentrations of methane and oxygen were approximately 20 mg L and 32 mg L, respectively.

Design Considerations. The effectiveness of *in situ* bioremediation is influenced by many factors, including: microorganisms, soils, oxygen, pH, temperature, type and quantity of contaminants, and nutrients.

Microorganisms. A large number of naturally occurring bacteria and fungi can use hydrocarbons for growth. The most commonly found genera of bacteria and fungi that are capable of hydrocarbon biodegradation are (1) bacteria: *Pseudomonas*, *Arthrobacter*, *Alkaligenes*, *Corynebacterium*, *Flavobacterium*, *Achromobacter*, *Micrococcus*, *Nocardia*, and *Mycobacterium* and (2) fungi: *Thricoderma*, *Penicillium*, *Aspergillus*, and *Mortierella* (19).

Soils contain a diverse culture of microorganisms, with the number and type present being a function of local environmental conditions. Usually, the limiting factor on the microbial population at an uncontaminated site is the amount of type of carbonaceous material available for biodegradation (19). However, there are almost always some microorganisms present in every soil environment that will take advantage of a new or different carbon source when it becomes available and will rapidly increase their populations to metabolize it. Addition of more or especially cultured microbes has been studied and applied in the field, however, some people believe that all that is needed in most cases is simply to optimize environmental conditions for the naturally occurring microbes. But for recalcitrant compounds or for more rapid biodegradation, specialized microbes may have potential. For example, the U.S. EPA has agreed to license specialized microbes that degrade chlorinated aromatic and chlorinated aliphatic compounds such as TCE under aerobic conditions which avoids the generation of vinyl chloride as a byproduct (20).

Soils. Soils with hydraulic conductivities greater than 10⁻⁴ centimeters per second (cm s) are good candidates for *in situ* bioremediation because they are permeable enough to allow the transport of oxygen and nutrients through the aquifer. The degree of homogeneity of the soil is important in order to predict ground water flow patterns for plume containment and distribution of oxygen and nutrients.

Contaminants. The type and concentration of contaminants in an aquifer dictate what type of *in situ* bioremediation system, aerobic, anaerobic, or combination, are the most effective.

Oxygen. Many microorganisms require either free oxygen or inorganic oxygen for growth. Some bacteria can use either free oxygen or inorganic oxygen for respiration, thus, these are called facultative bacteria. Many bacteria, however, require free oxygen for growth.

A rule of thumb used in wastewater treatment is that at least 1.5 mg L of dissolved oxygen is needed to prevent oxygen from being a growth limiting factor. Ref. 21 documents a threshold dissolved oxygen concentration of 0.5 mg L in ground water. Below 0.5 mg L, it was found that the biodegradation rates of benzene, toluene, and xylene were retarded. It was also shown that biodegradation rates increased until the ground water dissolved oxygen concentration reached 2 mg L at which point no further increase was observed. Thus, it appears that a target of 1.5 to 2.0 mg L dissolved oxygen maximizes biodegradation rates and put an upper limit on the cost to supply the oxygen.

Field data collected from ground water in spill areas have shown that dissolved oxygen concentrations are typically reduced below 1 mg L due to biodegradation. For example, at the site of a gasoline spill where gasoline-degrading bacteria had been identified, dissolved oxygen in the contaminated ground water ranged from 0 to 0.5 mg L while in the uncontaminated areas, levels were from 2.3 to 4 mg L (22). In another study of an aquifer contaminated with chlorinated solvents, dissolved oxygen was always less than 0.2 mg L (18). These studies demonstrate that biological growth and biodegradation of hydrocarbons is oxygen-limited, a fact which stimulates a continuing interest in developing effective oxygen sources and the means of distributing oxygen in ground water and soil.

Hydrogen Peroxide. Hydrogen peroxide is an effective means of getting more oxygen into the ground water for bioremediation and has many advantages over air or pure oxygen. Hydrogen peroxide has been demonstrated to not only boost dissolved oxygen levels, but it also stimulates microbial activity. It also has the advantage of chemically degrading contaminants partially or fully. Hydrogen peroxide mixes intimately with ground water for better distribution in the aquifer and can prevent injection well plugging caused by bacterial growth.

On the downside, various studies (23–25) have shown that hydrogen peroxide decomposes rapidly after soil contact, it is cytotoxic at a 3% solution and unless stabilized, oxygen bubbles can escape prematurely through the unsaturated zone before they have a chance to disperse well in the ground water.

Catalase enzymes generated by aerobic bacteria near the injection wells appear to be the main cause of hydrogen peroxide decomposition and oxygen off-gassing. Others have noted that hydrogen peroxide also reacts with dissolved iron, manganese, and other inorganics to produce oxides which can plug wells and formations (26,27). Precipitation of inorganic compounds during bioremediation has been studied in more detail (28). Concentrated solutions of hydrogen peroxide (30% w/v) used in bioremediation must be handled carefully to avoid skin burns.

Reports of ground water remediation studies using hydrogen peroxide are found frequently in the literature. Hydrogen peroxide reactions as an oxidant during bioremediation have been reviewed in Ref. 29. A field-scale study of an area contaminated by an aviation gasoline fuel spill showed that hydrogen peroxide did increase the concentration of oxygen downgradient of the injection well (24). Hydrogen peroxide was used to remediate ground water after a jet fuel leak at Eglin Air Force Base in Florida (25). Field applications using hydrogen peroxide for ground water cleanup are continuously reported by the U.S. EPA in its "Bioremediation in the Field" series. Treatment conditions common to these studies are the use of hydrogen peroxide and nutrient addition to stimulate biodegradation by the indigenous microorganisms.

Most reports citing hydrogen peroxide do not specify dosage rates. Concentrations as high as 500 mg/L have been shown to be nontoxic to most microorganisms (26). Other studies report dosage rates of 100 to 500 mg/L. Unpublished reports of rates as high as 1000 to 10,000 mg/L are not uncommon (24). Research studies have shown that a 0.05% solution (500 mg/L) was the maximum tolerated by a mixed culture of microorganisms capable of degrading gasoline. Concentrations up to 2000 mg/L could be tolerated if the dosage was stepped up incrementally; however, most toxicity studies showing no toxicity at relatively high concentrations are usually based on bacteria counts. Other studies have shown that concentrations greater than 100 mg/L decrease the *oxygen utilization rate* by microorganisms which is counterproductive to biodegradation, and therefore, increases the time needed for remediation.

Premature decomposition of hydrogen peroxide can be inhibited by adding monobasic potassium phosphate [7778-77-0] as a stabilizer (24). The phosphorus in the solution has the added benefit as a nutrient for the microorganisms. The concentration of 190 mg/L potassium phosphate was used in Huling's field study of ground water contaminated with aviation gasoline. The phosphate is added ahead of the hydrogen peroxide until breakthrough occurs in the downgradient wells. By adding the phosphate first, the phosphate binds with the inorganics in the soils and allows the hydrogen peroxide to be distributed further. Other chemicals have been used or studied to increase the stability of hydrogen peroxide are polyphosphates, stannate or phosphate, sodium pyrophosphate, citrate, and sodium silicate [1344-09-8] (30).

Hydrogen peroxide is also used as a treatment to clean wells of biological growth. One treatment is to pour a 0.5% solution into the well and allow it to sit for four hours after which pumping is resumed (31). Some treatments use solution concentrations as high as 10% to remove biofouling from wells (16).

Pure Oxygen. By increasing the oxygen level in soils and ground water, pure oxygen or oxygen-enriched streams stimulate bacterial activity. Using bottled oxygen, 35 to 40 mg/L of dissolved oxygen can be supplied at most field temperatures (31), which is less stable than hydrogen peroxide, a disadvantage with oxygen is that it may not distribute as well in the aquifer, although surfactants can promote highly stable microbubbles that are better able to travel throughout the ground water. Unlike hydrogen peroxide, oxygen does not aid in keeping the injection well free from biological growth that can plug the well.

Nitrates and Sulfates. There is much less information on actual applications of nitrate or sulfates as oxygen sources. Advantages of nitrate are that it is more soluble than oxygen, less costly, and less toxic than hydrogen peroxide. Regulatory agencies are less inclined to approve the use of nitrates in ground water cleanup as an oxygen or nutrient supplement because it can potentially contaminate drinking water supplies and plug the aquifer with biological growth. If nitrate is used, dosages must be carefully calculated and controlled to avoid these problems. In one study (32), nitrate was used as the oxygen source for microbial respiration in the cleanup of a drinking water aquifer contaminated with jet fuel and concentrations of benzene, toluene, ethylbenzene, and xylenes were treated below 5 µg/L. There is less literature on sulfate as an oxygen source than nitrate. Most references merely state that sulfate can be used as an oxygen source for anaerobic respiration.

Solid Phase Oxygen. Solid phase oxygen in the form of peroxides (calcium peroxide [1305-79-9], magnesium peroxide [1335-26-8]) has been investigated as a slow, continuous release system that avoids problems associated with the transient nature of molecular oxygen and hydrogen peroxide (33,34). Calcium peroxide releases oxygen into the system by first breaking down into hydrogen peroxide and calcium hydroxide; the hydrogen peroxide then decomposes to oxygen and water. Cages or socks of the material are suspended directly in the ground water well and the material can be blended into the soil for *ex situ* treatment and landfarming. For ground water treatment, this form of oxygen has the most potential for containment or remediation of low concentrations of dissolved organics.

Iron Oxides. Chelated iron oxide, using nitrilotriacetic acid [139-13-9] and EDTA, has been studied as an alternative oxygen source (35). Iron oxide which is often difficult for the microbes to access, is made more available by chelating agents.

Nutrients. In addition to carbon and oxygen, other nutrients are needed for microbial growth. Nitrogen, phosphorus, potassium [7440-09-7], calcium [7440-70-2], magnesium [7439-95-4], sodium [7440-23-5], sulfur [7704-34-9], chlorine, iron, manganese, cobalt, copper, boron [7440-42-8], zinc, molybdenum [7439-98-7], and aluminum [7429-90-5] are all present in the microbial cellular material (36). With the possible exceptions of nitrogen and phosphorus, which are present in the cell in relatively large amounts, these nutrients are needed at only trace levels and there is almost always enough in the natural environment to prevent growth-limiting conditions.

In the presence of large amounts of degradable carbon, the naturally available nitrogen and phosphorus could potentially limit growth and thus biodegradation rates. Based on a literature review (19) typical aerobic soil microorganisms contain 5 to 15 parts of carbon per part of nitrogen, with 10 parts being an acceptable average. This is about twice the 5:1 ratio of carbon to nitrogen that is usually assumed to represent the composition of the biomass in wastewater treatment processes (37). If this difference is real, it may be due to the specific environmental conditions in a particular location, since such conditions are known to influence the amount of nitrogen accumulated with cellular material (37). Phosphorus requirements are about one-fifth that of nitrogen (19,37).

These estimates of nitrogen and phosphorus do not consider that which is reintroduced by biological recycling. Bacterial growth and death continually recycle some fraction of the synthesized nitrogen and phosphorus. Ref. 19 cites an optimum carbon to nitrogen to phosphorus ratio of 250:10:3 for biodegradation of carbon compounds in soil, which accounts for the carbon oxidized to carbon dioxide and that assimilated into cellular material. Assuming that one part of carbon is incorporated into cellular material for every two parts of carbon that are used for energy and are converted to carbon dioxide, the ratio of carbon to nitrogen in the cell is about 8.3 to 1, a reasonably conservative estimate. A C:N:P ratio of 300:10:1 should be adequate for biodegradation, assuming no recycle (19).

Nutrients are usually added at concentrations ranging from 0.005 to 0.02% by weight (16). In a field application using hydrogen peroxide, nutrients were added to the injected water at the following concentrations: 380 mg/L ammonium chloride; 190 mg/L disodium phosphate, and 190 mg/L potassium phosphate, the latter used primarily to complex with iron in the formation to prevent decomposition of hydrogen peroxide (24).

Temperature. Ambient temperature is a key factor to successful biodegradation. In the range of 15 to 35°C, bacterial growth rates double for every 10°C rise in temperature. For most soil microorganisms, the optimal temperature for maximum growth is 30–35°C (19). Since vadose zone and ground water temperatures are typically in the range of 10 to 25°C (38), microbial growth will not be prevented, but will be at less than the maximum possible rates.

pH. A pH between 6 and 8 should be maintained for *in situ* bioremediation. Since most soil pHs are within this range, pH should not be a controlling factor for biodegradation of hydrocarbons, unless the soil has also been contaminated with acidic or caustic substances.

Air Sparging. With this technique, air is injected in the saturated zone to volatilize organics and to deliver oxygen for biodegradation. Soil vapor extraction may be used in conjunction with air sparging to capture the volatilized organics. Air sparging is used frequently at sites that have been contaminated with volatile compounds such as benzene, toluene, ethylbenzene, and xylenes (BTEx), common hazardous components of gasoline and other fuels. However, performance data are lacking and difficult to measure.

Biosparging. Biosparging is a form of air sparging, but the difference is that the primary purpose of biosparging is to deliver just enough air to meet oxygen requirements for bioremediation. Volatilization of organics may be an added benefit, but it is secondary to oxygen delivery. An enhancement is to sparge ozone in place of air. Besides providing oxygen for biodegradation, ozone aids in breaking down recalcitrant compounds such as chlorinated

ethenes, polycyclic aromatic hydrocarbons, and pentachlorophenol [87-86-5].

Biofilters. Biofilters, also known as biobarriers or microbial fences, are used to hinder migration of a contaminant plume. A biofilter is essentially a zone of biological activity that treats the contaminant as the ground water flows through the area. The biofilter zone can be established by installing a line of air sparging wells perpendicular to the direction of ground water flow.

Bioventing. Bioventing is soil venting that enhances biodegradation while extracting volatile compounds from the unsaturated zone.

Funnel-and-Gate. This *in situ* bioremediation method is so-named because ground water is funneled through openings or gates in an impermeable sheet piling or slurry wall. Zones of biological activity are created at the gates through air sparging. Contaminants are biodegraded as ground water is forced to pass through the gates. The funnel-and-gate method (also known as flume-and-gate) is used where ground water gradients are small. Because smaller treatment zones are created, operating expenses can be reduced and there is better process control.

Phytoremediation. Phytoremediation is a developing technology that uses plants, trees, and grasses to biodegrade, extract, or stabilize organic and metal contaminants in soil and ground water. Engineered wetlands is a well-known form of phytoremediation. Phytoremediation is likely to be used in conjunction with other treatment processes, primarily as a polishing step (39). As reported, this method has potential in the cleanup of residual contamination in soil micropores, hydraulic control (poplars and weeping willows pump 190 to 1325 L per day per tree, alternative landfill caps for leachate control, leachate treatment, and buffer zones for plume containment.

Ex Situ Bioremediation.

Biological Wastewater Treatment. Ground water and leachate can be treated in a biological wastewater treatment system such as rotating biological contactors (RBCs), trickling filters, sequencing batch reactors, fluidized bed reactors, activated sludge, or aerated impoundments. For biological treatment to be feasible, the ground water or leachate must contain sufficient organics to support microbial growth; otherwise, dilute streams should be treated by physical–chemical means or combined with higher strength wastewaters for biological treatment in a larger wastewater treatment system (for example, in an onsite wastewater plant treating process wastewaters at a manufacturing facility). A sequencing batch reactor is similar to activated sludge except that treatment takes place in one tank in batch mode, ie, biodegradation with mixing, followed by quiescent settling and sludge withdrawal. A fluidized bed reactor uses granular activated carbon or sand as the support media for microbial growth. The ground water to be treated is introduced through the bottom of the reactor which serves to fluidize or suspend the support media. Pretreatment to remove solids is required to avoid plugging the reactor.

Slurry-Phase Soil Treatment. Contaminated soil can be treated biologically in slurry form, not unlike biological treatment of wastewater. However, one important difference with contaminated soil is that the contaminants preferentially adsorb unto the soil particles and must be desorbed before the microorganisms can effectively degrade the compounds. Contaminated soil may be pretreated prior to the bioreactor, for example, by soil washing to separate the fine particles which normally contain the majority of the contamination. Contaminated soil can be slurried by mixing with water, wastewater, or contaminated ground water. The solids concentration of the slurry can be fairly high, upwards to 50% , which makes mixing to keep the soil in suspension an important design parameter.

Soil Heaps, Piles, Beds, and Windrows. Contaminated soil can be excavated and placed in heaps, piles, beds, or windrows. Other organic or bulking agents such as wood chips or straw, may be added to aid in composting, mixing, and aerating the soil. In a windrow system, the soil is placed in long rows and turned periodically to mix and aerate the soil. Soil piles and heaps differ in size, heaps being larger and used where large volumes of contaminated soil must be treated. Piles and heaps may be covered or uncovered, depending on the volatility of the contaminants and air emissions requirements and whether a vacuum is drawn through or air is blown into the soil. With treatment beds, contaminated soil is placed on top of a liner system consisting of a synthetic liner, sand, and leachate drainage pipe. The bed may be covered or enclosed. In any of these treatment systems, volatilized contaminants that are removed by vacuum may be treated by incineration, carbon adsorption, or vapor-phase biodegradation. Leachate may be treated biologically in a wastewater treatment system.

Physical–Chemical–Thermal *In Situ* Treatment Related Specifically to Soils and Ground Water

Electrokinetics. Electrokinetics is a tested technology that has been used for over half a century to dewater and stabilize soils, and has recently been investigated for in situ use at hazardous waste sites (23). Primarily used for metals removal, the technology utilizes an electrical field to generate a flow and concentration gradient in porous and semiporous soils.

An innovative technology called the "lasagna" process is based on the electrokinetic phenomenon called electroosmosis. The lasagna process was created to treat difficult wastes in low permeability, silt- and clay-laden soils (40). The lasagna process is so named because it consists of a number of layered subsurface electrodes and treatment zones. These layers can be constructed either horizontally where contaminants are forced to move upward or in vertical position where lateral contaminant movement is desired.

A low voltage electric current is applied by subsurface electrodes which forces the migration of contaminants from low permeable soils into high permeable treatment zones. These treatment zones that may be created by hydrofracturing (injecting a slurry to create porous pancake-shaped zones), directional drilling, or sheet piling. Treatment in these zones may include physical, chemical, or biological processes such as adsorption by activated carbon which would then serve as a biodegradation matrix.

In Situ Air Stripping. An innovation to conventional pump and treat air stripping is *in situ* air stripping. Two horizontal wells are installed, one below the water table and one in the vadose zone. Air is injected in the lower well while contaminated soil vapor is extracted by vacuum through the upper well.

Soil Flushing. Soil flushing is similar to soil washing of excavated soils except flushing is done *in situ* below the ground surface. With soil flushing, the contaminated area is flooded with water. Surfactants or detergents may be added to the water to enhance removal of organics.

Soil Vapor Extraction. Volatile compounds can be extracted from subsurface soils by applying a vacuum. An added benefit is that the vacuum pulls in air from the ground surface that stimulates biodegradation through increased oxygen transport. The ground water below the vadose zone can also be aerated to volatilize contaminants in the ground water (air sparging) which then move into the vadose zone where they are removed along with the soil vapors.

A U.S. EPA study (41) showed that soil vapor extraction (SVE) is an effective treatment for removing volatile contaminants from the vadose zone. Sandy soils are more effectively treated than clay or soils with higher organic content because higher air flows are possible in sand and clays–organic soils tend to adsorb or retain more contaminants. Removal of volatiles is rapid in the initial phase of treatment and thereafter decreases rapidly thereafter-an important consideration in the design of air emissions control over the life of the project.

An innovative companion technology to SVE is radio frequency heating of the soil (42,43). Heating the soil increases the volatilization of the containments which are removed by SVE. Antennae are installed near the center of the contaminated area; the radio frequency energy applied through the antennae heat the soil to target levels of 100 to 150°C.

Vitrification. Vitrification is an innovative treatment that turns contaminated soils into a glasslike monolithic mass. Heat is applied through electrodes placed in the ground, the soil reaching temperatures of 1600 to 2000°C. A layer of graphite [7782-42-5] and glass frit is first placed on the surface of the ground between the electrodes to act as the initial conductive starter path. The conductive layer and adjacent soils become a molten mass that becomes the primary electrical conductor and heat transfer medium. As heat continues to be supplied by the electrodes, the molten mass moves both outward and downward. As organic contaminants in the soil are heated, they begin to vaporize and eventually pyrolyze with the rising temperature. Inorganic contaminants are immobilized in the molten material. Off-gases, which include vaporized organics and the by-products of organic and inorganic pyrolysis, are captured above the site and treated to meet air emissions standards. Once cooled, the vitrified mass is very stable with low leaching potential.

Vitrification is effective at destroying and immobilizing hazardous materials, but it is very energy intensive and thus expensive. Consequently, it is used primarily where wastes are difficult to treat or destruction–immobilization of contaminants is very important such as with radionuclides.

Pump and Treat

In the early years of ground water and soil remediation, pump and treat was the conventional technology. Contaminated ground water is pumped to the surface where it is treated and reinjected or discharged to surface waters or wastewater treatment plants. Reinjection may be used to stimulate *in situ*

bioremediation, where the treated ground water is enriched with oxygen and nutrients prior to reinjection. Many years of experience with pump and treat methods have brought to light certain limitations, as well as improvements and enhancements. Therefore, pump and treat technology must be reviewed for its applicability to a particular site.

Pump and treat technology is inherently slow because it depends on ground water for transport of the contaminant to the extraction well. This characteristic is particularly troublesome when the contaminant is only slightly water soluble, adheres to the soil, or collects in pools within the aquifer.

There are many cases of contamination by dense nonaqueous phase liquids (DNAPLs) that have frustrated pump and treat efforts. The general consensus is that pump and treat can reduce contamination or keep it from spreading, but it has failed in many cases to remediate aquifers to stringent cleanup goals.

Common technologies for surface treatment of contaminated ground water are air stripping, carbon adsorption, biodegradation, and ion exchange. If cleanup goals are in the µg L range, then typically pump and treat options are limited to air stripping and carbon adsorption. Conventional biological treatment is usually not suitable for the relatively low levels of contaminants found in ground water. Biological treatment of extracted ground water is discussed under "Bioremediation." Air stripping, carbon adsorption, and ion exchange are discussed under "Physical Chemical Treatment Technologies."

Extraction Techniques

Two-phase vacuum extraction (TPVE) is a way of pumping both contaminated air and ground water through a single well. A vacuum is applied through a well that is screened both through the vadose and ground water zones. Both air and ground water are pumped to the surface where they are separated for treatment. Advantages of this system are that it does not require separate wells for air extraction and ground water pumping and that water pumping rates can be greatly increased.

Ex Situ Soil Nonbiological Treatment

Soil Leaching. Soil leaching or acid extraction uses acid to solubilize metals for removal from soils, a technique akin to that in the mining industry. After extraction with an acid such as hydrochloric, sulfuric, or nitric, the soil is separated from the acid, rinsed with water to remove excess acid and metals, dewatered, and neutralized. The acid is regenerated and recycled back to the process. The extracted metals can be precipitated and recovered.

Soil Washing. Soil washing is as simple a process as the name implies. Soil is excavated, physically broken up, washed with a water-based solution, and separated into size fractions (sand–gravel and silt–clay). The goal of soil washing is to transfer pollutants to the wash water and to isolate the finer soil fraction that preferentially adsorbs organics and metals. Various surfactants, extractants, or detergents are added to the wash water to enhance removal of organic contaminants from the soil particles. Soil fines and contaminants are removed from the wash water which is then recycled back to the process.

The finer soil fraction contains adsorbed organics, small metallic particles, and bound ionic metals. This fraction may be treated further to remove the contaminants, or it may be incinerated or landfilled. The "clean" coarse fraction may contain some residual metallic fragments. With metal contamination, both the fine and coarse soil fractions may be leached with an acid solution to remove the metals.

Other Techniques. Other methods, more conventional in type, are employed for *ex-situ* treatment. These include solvent extraction and thermal desorption, which are detailed under the "Physical Chemical Treatment" and "Thermal Treatment" sections, respectively.

BIBLIOGRAPHY

1.

Water Environment Federation (formerly Water Pollution Control Federation), *Hazardous Waste Treatment Processes*, Manual of Practice FD-18, prepared by Task Force on Hazardous Waste Treatment, Alexandria, Va., 1990.

2.

D. R. Roeck and A. Sigg, *Environ. Protection*, **7**(1) (1996).

3.

J. L. Musterman and V. J. Boero, *Industrial Wastewater* **3**(2), 49–55 (1995).

4.

U.S. Environmental Protection Agency, *SITE Demonstration Bulletin: ZenoGem™ Wastewater Treatment Process*, ZENON Environmental Systems, EPA 540 MR-95 503, Cincinnati, Ohio, 1995.

5.

U.S. Environmental Protection Agency, *SITE Emerging Technology Summary: Cross-Flow Pervaporation for Removal of VOCs from Contaminated Wastewater*, EPA 540 SR-94 512, Cincinnati, Ohio, 1994.

6.

W. J. Weber, Jr., *Physiochemical Processes for Water Quality Control*, John Wiley & Sons, Inc., New York, 1972.

7.

U.S. Environmental Protection Agency, *Emerging Technology Summary: Laser Induced Photochemical Oxidative Destruction of Toxic Organics in Leachates and Groundwaters*, EPA 540 SR-92 080, Cincinnati, Ohio, 1993.

8.

U.S. Environmental Protection Agency, *Demonstration Bulletin: CAV-OX® Ultraviolet Oxidation Process*, Magnum Water Technology, EPA 540 MR-93 520, Cincinnati, Ohio, 1993.

9.

U.S. Environmental Protection Agency, *perox-pure Chemical Oxidation Technology Peroxidation Systems, Inc., Applications Analysis Report*, EPA 540 AR-93 501, Washington, D.C., 1993.

10.

U.S. Environmental Protection Agency, *CAV-OX Cavitation Oxidation Process*, Magnum Water Technology, Inc., *Application Analysis Report*, EPA 540 SR-93 520, 1994.

11.

R. Jensen, *Environmental Protection*, **5**(6), 16–17 (1994).

12.

P. Pitter and J. Chudoba, *Biodegradability of Organic Substances in the Aquatic Environment*, CRC Press, Boca Raton, Fla., 1990.

13.

W. W. Eckenfelder, Jr. and J. L. Musterman, *Activated Sludge Treatment of Industrial Wastewater*, Technomic Publishing Company, Lancaster, Pa., 1995.

14.

R. Noyes, *Unit Operations in Environmental Engineering*, Noyes Publications, Park Ridge, N.J., 1994.

15.

P. N. Cheremisinoff, *The Nat. Environ. J.* **4**(1), 49–54 (Jan. Feb. 1994).

16.

J. L. Sims and co-workers, *In Situ Bioremediation of Contaminated Ground Water*, U.S. Environmental Protection Agency, EPA 540 S-92 003, Washington, D.C., 1992.

17.

A. Hasbach, *Pollution Eng.* **25**(6) (1993).

18.

L. Sempriani and co-workers, *A Field Evaluation of In Situ Biodegradation for Aquifer Restoration*, U.S. Environmental Protection Agency, EPA 600 2-87 096, Washington, D.C., 1987.

19.

R. N. Miller, "A Field-Scale Investigation of Enhanced Petroleum Hydrocarbon Biodegradation in the Vadose Zone Combining Soil Venting as an Oxygen Source with Moisture and Nutrient Addition," doctoral dissertation submitted to the Civil and Environmental Engineering Department, Utah State University, Logan, Utah, 1990.

20.

U.S. Environmental Protection Agency, "EPA Signs License Agreement for TCE-Degrading Microorganisms," in *Bioremediation in the Field*, Center for Environmental Research Information, EPA 540 N-93 001, Cincinnati, Ohio, no. 8, 1993.

21.

M. W. Kemblowski and co-workers, "Fate and Transport of Residual Hydrocarbon in Ground Water: A Case Study," Petroleum Hydrocarbons and Organic Chemicals in Ground Water: Prevention, Detection, and Restoration, presented at the conference and exposition, National Water Well Association and American Petroleum Institute, Nov. 17–19, 1987.

22.

R. L. Raymond and co-workers, *Field Report, Field Application of Subsurface Biodegradation of Gasoline in a Sand Formation*, Project No. 307-77, Committee on Environmental Affairs, American Petroleum Institute, Washington, D.C., 1976.

23.

C. D. Chambers and co-workers, *In Situ Treatment of Hazardous Waste-Contaminated Soils*, Noyes Data Corporation, Park Ridge, N.J., 1991.

24.

S. G. Huling and co-workers, *Enhanced Bioremediation Utilizing Hydrogen Peroxide as a Supplemental Source of Oxygen: A Laboratory and Field Study*,

EPA 600 2-90 006, U.S. Environmental Protection Agency, Washington, D.C., 1990.

25. T. Pheiffer, "Case Study of the Eglin Air Force Base Ground Water Bioremediation System," in *Innovative Operational Treatment Technologies for Application to Superfund Sites: Nine Case Studies*, U.S. Environmental Protection Agency, Washington, D.C., 1990.

26. D. L. Russell, *Remediation Manual for Petroleum-Contaminated Sites*, Technomic Publishing Co., Inc., Lancaster, Pa., 1992.

27. P. J. Ware, *The Nat. Environ. J.*, **3**(4) (1993).

28. P. Morgan and R. J. Watkinson, *Water Research* **26** (1992).

29. D. L. Pardieck and co-workers, *J. Contam. Hydrol.* **9** (1992).

30. K. Elizardo, *Pollution Eng.* **25**(11), 38–39 (1993).

31. R. B. King and co-workers, *Practical Environmental Bioremediation*, Lewis Publishers, Boca Raton, Fla., 1992.

32. Hutchins and co-workers, *Nitrate for Bioremediation of an Aquifer Contaminated with Jet Fuel, Report for June 1988–Sept. 1990*, U.S. Environmental Protection Agency in cooperation with Solar Universal Technologies, Inc., NSI Technology Services Corp., Traverse Group, Inc., and 9th Coast Guard District, 1991.

33. G. R. Brubaker, *Civil Eng.* **65**(10), 38–41 (1995).

34. M. R. Fagan, *Environmental Protection* **5**(9), 45–52 (1994).

35. Water Environment Federation, *Water Environ. Technol.* **6**(8), 26 (1994).

36. C. E. Clifton, *Introduction to Bacterial Physiology*, McGraw-Hill Book Company, Inc., New York, 1957.

37. W. W. Eckenfelder, Jr., *Industrial Water Pollution Control*, McGraw-Hill Book Company, New York, 1966.

38. D. P. Hart and R. Couvillion, *Earth-Coupled Heat Transfer*, National Water Well Association, Dublin, Ohio, 1986.

39. K. Matso, *Civil Eng.* **65**(10), 46–49 (1995).

40. U.S. Environmental Protection Agency, "EPA and Industry Sign CRADA to Develop Innovative 'Lasagna' Process," in *Bioremediation in the Field*, Center for Environmental Research Information, Cincinnati, Ohio, 1994.

41. M. H. Corbin, N. A. Metzger, and M. F. Kress, *Project Summary: Field Investigation of Effectiveness of Soil Vapor Extraction Technology*, EPA 600 SR-94 142, U.S. Environmental Protection Agency, Cincinnati, Ohio, 1994.

42. U.S. Environmental Protection Agency, *SITE Technology Capsule: IITRI Radio Frequency Heating Technology*, EPA 540 R-94 527a, Cincinnati, Ohio, 1995.

43. U.S. Environmental Protection Agency, *SITE Technology Capsule: KAI Radio Frequency Heating Technology*, EPA 540 R-94 528a, Cincinnati, Ohio, 1995.

General References

C. E. Adams, Jr., D. L. Ford, and W. Wesley Eckenfelder, Jr., *Development of Design and Operational Criteria for Wastewater Treatment*, Enviro Press, Inc., Nashville, Tenn., 1981.

API, *In Situ Air Sparging: Evaluation of Petroleum Industry Sites and Considerations for Applicability, Design, and Operation*, Publication No. 841-46090, American Petroleum Institute, Washington, D.C., 1995.

API, *Modeling Aerobic Biodegradation of Dissolved Hydrocarbons in Heterogenous Geologic Formations*, Publication No. 848-00200, American Petroleum Institute, Washington, D.C., 1995.

K. W. Benker, "Removing Metals from Soil," *Civil Eng.* **65**(10) (1995).

E. J. Bouwer and G. D. Cobb, *In Situ Groundwater Treatment Technology Using Biodegradation*, U.S. Army Toxic and Hazardous Materials Agency, Report AMXTH-TE-CR-88023, Washington, D.C., 1987.

G. R. Chaudhry, ed., *Biological Degradation and Bioremediation of Toxic Chemicals*, Dioscorides Press, Portland, Oreg., 1994.

N. F. Gray, *Activated Sludge: Theory and Practice*, Oxford Science Publications, New York, 1990.

D. Jenkins, M. G. Richard, and G. T. Daigger, *Manual on the Causes and Control of Activated Sludge Bulking and Foaming*, Lewis Publishers, Boca Raton, Fla., 1993.

E. R. Krishnan and co-workers, *Recovery of Metals from Sludges and Wastewaters*, Noyes Data Corp., Park Ridge, N.J., 1993.

M. D. LaGrega, P. L. Buckingham, and J. C. Evans, *Hazardous Waste Management*, McGraw-Hill, Inc., New York, 1994.

National Research Council, *Alternatives for Ground Water Cleanup*, National Academy Press, Washington, D.C., 1994.

National Research Council, *In Situ Bioremediation: When Does It Work?*, National Academy Press, Washington, D.C., 1993.

U.S. EPA, *Compendium of Technologies Used in the Treatment of Hazardous Wastes*, EPA 625 8-87 014, Cincinnati, Ohio, 1987.

U.S. EPA, *Assessing UST Corrective Action Technologies: Early Screening of Cleanup Technologies for the Saturated Zone*, EPA 600 2-90 027, prepared by P. J. Reidy and co-workers, U.S. Environmental Protection Agency, Washington, D.C., 1990.

U.S. EPA, *Anaerobic In Situ Treatment of Chlorinated Ethenes*, EPA 600 D-90 204, prepared by G. W. Sewell and co-workers, U.S. Environmental Protection Agency, Washington, D.C., 1990.

U.S. EPA, "Case Study of Ground Water Extraction with Air Stripping, Des Moines TCE Site," in *Innovative Operational Treatment Technologies for Application to Superfund Sites: Nine Case Studies*, prepared by T. Pheiffer, U.S. Environmental Protection Agency, Washington, D.C., 1990.

U.S. EPA, "Case Study of Ground Water Extraction with Air Stripping, Eau Claire Municipal Well Field Site," in *Innovative Operational Treatment Technologies for Application to Superfund Sites: Nine Case Studies*, prepared by T. Pheiffer, U.S. Environmental Protection Agency, Washington, D.C., 1990.

U.S. EPA, *Treatment Technology Background Document*, PB91-160556, Washington, D.C., 1991.

U.S. EPA, *Low Temperature Thermal Treatment (LT₃®) Technology*, Roy F. Weston, Inc., Applications Analysis Report," EPA 540 AR-92 019, Washington, D.C., 1992.

U.S. EPA, *TCE Removal from Contaminated Soil and Ground Water*, Ground Water Issue, EPA 540 S-92 002, prepared by H. H. Russell and co-workers, U.S. Environmental Protection Agency, Washington, D.C., 1992.

U.S. EPA, *Mercury and Arsenic Wastes: Removal, Recovery, Treatment, and Disposal*, Noyes Data Corp., Park Ridge, N.J., 1992.

U.S. EPA, *Silicate Technology Corporation's Solidification/ Stabilization Technology for Organic and Inorganic Contaminants in Soils: Applications Analysis Report*, EPA 540 AR-92 010, Washington, D.C., 1992.

U.S. EPA, *Membrane Treatment of Wood Preserving Site Groundwater by SBP Technologies, Inc., Applications Analysis Report*, EPA 540 AR-92 014, Washington, D.C., 1993.

U.S. EPA, *Resources Conservation Company B.E.S.T® Solvent Extraction Technology: Applications Analysis Report*, EPA 540 AR-92 079 (1993).

U.S. EPA, *SITE Demonstration Bulletin: Microfiltration Technology*, EPOC Water, Inc., EPA 540 MR-93 513, Cincinnati, Ohio, 1993.

U.S. EPA, *Eco Logic International Gas-Phase Chemical Reduction Process, The Thermal Desorption Unit: Applications Analysis Report*, EPA 540 AR-94 504, Washington, D.C., 1994.

U.S. EPA, *Emerging Technology Summary: Acid Extraction Treatment System for Treatment of Metal Contaminated Soils*, EPA 540 SR-94 513, prepared by S. W. Paff, B. E. Bosilovich, and N. J. Kardos, U.S. Environmental Protection Agency, Cincinnati, Ohio, 1994.

U.S. EPA, *SITE Demonstration Bulletin, Thermal Desorption System: Clean Berkshires, Inc.*, EPA 540 MR-94 507, Cincinnati, Ohio, 1994.

U.S. EPA, *Emerging Technology Bulletin: Volatile Organic Compound Removal from Air Streams by Membranes Separation, Membrane Technology and Research, Inc.*, EPA 540 F-94 503, Cincinnati, Ohio, 1994.

U.S. EPA, *SITE Technology Capsule: Clean Berkshires, Inc. Thermal Desorption System*, EPA 540 R-94 507a, Cincinnati, Ohio, 1994.

U.S. EPA, *Geosafe Corporation In Situ Vittrification: Innovative Technology Evaluation Report*, EPA 540 R-94 520, Washington, D.C., 1995.

U.S. EPA, *Manual: Ground Water and Leachate Treatment Systems*, EPA 625 R-94 005, Washington, D.C., 1995.

U.S. EPA, *Project Summary: Assessing UST Corrective Action Technologies: Lessons Learned about In Situ Sparging at the Denison Avenue Site, Cleveland, Ohio*, EPA 600 SR-95 040, Cincinnati, Ohio, 1995.

U.S. EPA, *Project Summary: Removal of PCBs from Contaminated Soil Using the CF Systems® Solvent Extraction Process: A Treatability Study*, EPA 540 SR-95 505, Cincinnati, Ohio, 1995.

U.S. EPA, *SITE Technology Capsule: Terra-Kleen Solvent Extraction Technology*, EPA 540 R-94 521a, Cincinnati, Ohio, 1995.

Water Environment Federation (formerly Water Pollution Control Federation) and American Society of Civil Engineers, *Aeration: A Wastewater Treatment Process*, Manual of Practice No. FD-13, Alexandria, Va. and New York (1988).

Woodside, *Hazardous Materials and Hazardous Waste Management: A Technical Guide*, John Wiley & Sons, Inc., New York, 1993.

Dianna S. Kocurek
Tischler/Kocurek

Gayle Woodside
IBM Corporation

WASTES, INDUSTRIAL

Recent legislation in the United States (ca 1997) has added several new parameters to the requirements for effluent permits. Parameters to be considered now, depending on the plant location, are biochemical oxygen demand (BOD), total suspended solids (TSS), chemical oxygen demand (COD), volatile organic compounds (VOC), priority pollutants, aquatic toxicity, heavy metals, nitrogen, and phosphorus. Because most industries have an existing wastewater treatment plant, in order to meet these new criteria upgrading, pretreatment, or tertiary treatment will be required. Innovative technologies and modifications to existing technologies are now (ca 1997) becoming available.

Waste Minimization

Before end-of-pipe wastewater treatment or modifications to existing wastewater treatment facilities to meet new effluent criteria are undertaken, a program of waste minimization should be initiated.

Reduction and recycling of waste are inevitably site- and plant-specific, but a number of generic approaches and techniques have been used successfully across the United States to reduce many kinds of industrial wastewaters.

Generally, waste minimization techniques can be grouped into four major categories: inventory management and improved operations, modification of equipment, production process changes, and recycling and reuse. Such techniques can have applications across a range of industries and manufacturing processes, and can apply to hazardous as well as nonhazardous wastes.

Many of these techniques involve source reduction, the preferred option in the U.S. Environmental Protection Agency's (EPA's) hierarchy of waste management techniques. Others deal with on- and off-site recycling. The best way to determine how these general approaches can be designed to fit a particular company's needs is to conduct a waste minimization assessment. In practice, waste minimization opportunities are limited only by the extent of the ingenuity of the generator. In the end, a company looking carefully at overall returns of waste minimization may well conclude that the most feasible strategy would be to use a combination of source reduction and recycling projects. Waste minimization approaches as developed by the EPA are shown in Table 1.

Table 1. Waste Minimization Approaches and Techniques

Approach	Related techniques
inventory management and improved operations	inventory and trace all raw materials
	purchase fewer toxic and more nontoxic production materials
modification of equipment	implement employee training and management feedback
	improve material receiving, storage, and handling practices
	install equipment that produces minimal or no waste
	modify equipment to enhance recovery or recycling options
	redesign equipment or production lines to produce less waste
production process changes	improve operating efficiency of equipment
	maintain strict preventive maintenance program
	substitute nonhazardous for hazardous raw materials
	segregate wastes by type for recovery
	eliminate sources of leaks and spills
recycling and reuse	separate hazardous from nonhazardous wastes
	redesign or reformulate end products to be less hazardous
	optimize reactions and raw material use
	install closed-loop systems
	recycle on-site for reuse
	recycle off-site for reuse
	exchange wastes

The six major ways of reducing pollution are as follows:

- (1) Recirculation. In the paper board industry, white water from a paper machine can be put through a saveall to remove the pulp and fiber and then be recycled to various points in the paper-making process.
- (2) Segregation. Clean streams are separated for direct discharge. Concentrated or toxic streams are separated for separate treatment.
- (3) Disposal. In many cases, concentrated wastes can be removed in a semidry state. In the production of ketchup, the kettle bottoms after cooking and preparation of the product are usually flushed to the sewer. The total discharge BOD and suspended solids can be markedly reduced by removal of this residue in a semidry state for disposal. In breweries, the secondary storage units have a sludge in the bottom of the vats that contains both BOD and suspended solids. Removal of this as a sludge rather than flushing to the sewer will reduce the organic and solids load to treatment.
- (4) Reduction. It is common practice in many industries, such as breweries and dairies, to have hoses continuously running for clean-up purposes. The use of automatic cutoffs can substantially reduce the wastewater volume.
- (5) The use of drip pans to catch products, in cases such as a dairy or ice-cream manufacturing plant, instead of flushing this material to the sewer, considerably reduces the organic load. A similar case exists in the plating industry where a drip pan placed between the plating bath and the rinse tanks will reduce the metal dragout.
- (6) Substitution. The substitution of chemical additives of a lower pollutional effect in processing operations, eg, substitution of surfactants for soaps in the textile industry.

Water reuse is usually a question of the tradeoff between the costs of raw water and the costs associated with treatment for reuse and for discharge. If biological treatment is to be employed, several factors must be considered. These are an increase in concentration of organics, both degradable and nondegradable. This may have a negative effect in terms of final effluent toxicity. An increase in temperature or total dissolved solids may adversely affect the performance of the biological process.

Characterization of Wastewaters

A comprehensive analytical program for characterizing wastewaters should be based on relevance to unit treatment process operations, the pollutant or pollutants to be removed in each, and effluent quality constraints. The qualitative and quantitative characteristics of waste streams to be treated not only serve as a basis for sizing system processes within the facility, but also indicate streams having refractory constituents, potential toxicants, or biostats. Such streams are not amenable to effective biological treatment, as indicated by the characterization results, and require treatment using alternative processes.

It should be recognized that the total volume of wastewater as well as the chemical analyses indicating the organic and inorganic components are required, backed by statistical validity, before the conceptualizing of the overall treatment plant design can begin. The basic parameters in wastewater characterization are summarized in Table 2.

Table 2. Basic Parameters in Wastewater Characterization ^a

Parameter	Examples
chemical composition	Basic parameters in wastewater characterization
	source information for the individual points of origin
	waste constituents (specific compounds or general composition)
	discharge rate (average and peak)
	batch discharges
	frequency of emergency discharges or spills
	organic and inorganic constituents
	gross organics
	chemical oxygen demand (COD)
	total organic carbon (TOC)
	biochemical oxygen demand (BOD)
	extractables
	toxics, hazardous compounds, priority pollutants
	gross inorganics—total dissolved solids
	specific inorganic ions (As, Ba, Cd, CN, Hg, Pb, Se, Ni, Sn, nitrates)
physical properties	pH, acidity, alkalinity
	nitrogen and phosphorus
	oil and grease
	oxidizing reducing agents, eg, sulfides
	surfactants
	chlorine demand
	temperature range and distribution
	particulates: colloidal, settleable, and floatable solids
	color
	odor
biological factors	foamability
	corrosiveness
	radioactivity
	biochemical oxygen demand
	toxicity (aquatic life, bacteria, animals, plants)
flow characteristics	pathogenic bacteria
	average daily flow rate
	duration and magnitude of peak flow rate
	maximum rate of change of flow rate
	storm-water flow rate (average and peak)
	<i>Causes of Variability in Waste Characterization</i>
	changes in production rate
	variations in plant product mix
	batch operations
	variations in efficiencies of production units
	changes in raw materials
	upsets in production processes
	maintenance (equipment shutdown and cleanout)
	miscellaneous leaks and spills
	contaminated drainage and runoff from rainstorms

^a Ref. 1

Industrial Wastewater Flow

The design flows for industrial complexes generally consist of the following: (1) base process flows resulting from normal production operations; (2) sanitary sewage; (3) contaminated storm runoff; (4) other sources, eg, extraordinary dumps, tank draining, and ballast discharge. The base flow and sanitary contribution can be measured in open channels or closed conduits using a variety of methods, such as automatic metering devices, weirs, or less sophisticated devices. Care should be taken to ensure flows are measured during workday and weekend operations, different work shifts, and over a sufficiently long period of time to reflect statistical reliability.

Since the mid-1980s, contaminated storm runoff has become an object of increasing concern within industrial complexes. Storm flow is intermittent and unpredictable in nature, and little data has been collected to typify its characteristics. The level of flow and degree of contamination not only varies within an installation; it has its own geometric characteristics, which influence patterns of surface runoff.

Definition of Wastewater Constituents

Parameters used to characterize wastewaters can be classified as organic and inorganic analyses. The organic content of wastewater is estimated in terms of oxygen demand using biochemical oxygen demand (BOD), chemical oxygen demand (COD), or total oxygen demand (TOD). Additionally, the organic fraction can be expressed in terms of carbon, using total organic carbon (TOC). It should be understood that these parameters do not necessarily measure the same constituents. Specifically, they reflect the following: (1) BOD: biodegradable organics in terms of oxygen demand; (2) COD: organics amenable to chemical oxidation as well as certain inorganics, such as sulfides, sulfites, ferrous iron, chlorides, and nitrites; (3) TOD: all organics and some inorganics in terms of oxygen demand; and (4) TOC: all organic carbon expressed as carbon.

It is important to identify volatile organic carbon (VOC) and the presence of specific priority pollutants, in addition to the total organic content. The

organic characteristics of various industrial wastewaters are shown in Table 3.

Table 3. Oxygen Demand and Organic Carbon of Industrial Wastewaters

Waste	BOD, mg L	COD, mg L	TOC, mg L	BOD TOC	COD TOC
chemical ^a		4,260	640		6.65
chemical ^a		2,410	370		6.60
chemical ^a		2,690	420		6.40
chemical		576	122		4.72
chemical	24,000	41,300	9,500	2.53	4.35
chemical–refinery		580	160		3.62
petrochemical		3,340	900		3.32
chemical	850	1,900	580	1.47	3.28
chemical	700	1,400	450	1.55	3.12
chemical	8,000	17,500	5,800	1.38	3.02
chemical	60,700	78,000	26,000	2.34	3.00
chemical	62,000	143,000	48,140	1.28	2.96
chemical		165,000	58,000		2.84
chemical	9,700	15,000	5,500	1.76	2.72
nylon polymer		23,400	8,800		2.70
petrochemical					2.70
nylon polymer		112,600	44,000		2.50
olefin processing		321	133		2.40
butadiene processing		359	156		2.30
chemical		350,000	160,000		2.19
synthetic rubber		192	110		1.75

^a High concentration of sulfides and thiosulfates.

The inorganic characterization schedule for wastewaters to be treated using biological systems should include those tests which provide information concerning (1) potential toxicity, such as heavy metal, ammonia, etc; (2) potential inhibitors, such as total dissolved solids (TDS) and chlorides; (3) contaminants requiring specific pretreatment such as pH, alkalinity, acidity, suspended solids, etc; and (4) nutrient availability.

Aquatic toxicity is becoming (ca 1997) a permit requirement on all discharges. Aquatic toxicity is generally reported as an LC₅₀ (the percentage of wastewater which causes the death of 50% of the test organisms in a specified period ie, 48 or 96 h, or as a no observed effect level (NOEL), in which the NOEL is the highest effluent concentration at which no unacceptable effect will occur, even at continuous exposure.

Toxicity is also frequently expressed as toxicity units (TU), which is 100 divided by the toxicity measured:

$$TU = \frac{100}{LC_{50} \text{ or NOEL}}$$

inwhich the LC₅₀ or the NOEL is expressed as the percent effluent in the receiving water. Therefore, an effluent having an LC₅₀ of 10% contains 10 toxic units.

Effluent toxicity can also be defined as a chronic toxicity in which the growth or reproduction rate of the species is affected.

Wastewater Treatment

In order to meet present (ca 1997) requirements, existing plants need to be retrofitted and new plants must incorporate advanced wastewater treatment technology. A substitution flow sheet showing available technology is shown in Figure 1. Options for meeting restrictive requirements are tertiary treatment following biological treatment, source treatment of toxic or refractory wastewaters, or modifications of the existing biological technology incorporating powdered activated carbon (PAC). These technologies are discussed herein in detail.

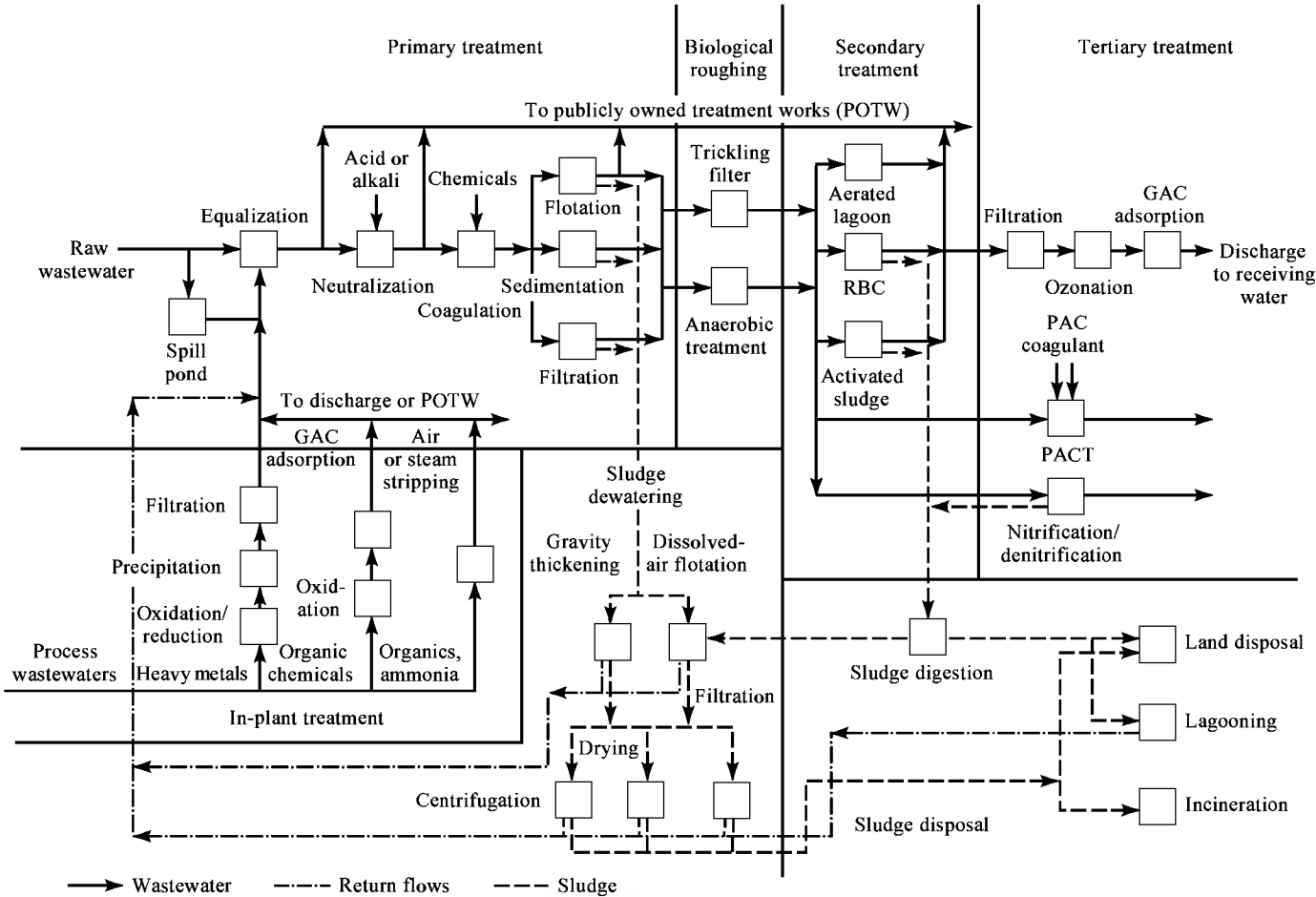


Fig. 1. Alternative wastewater treatment technologies, where GAC = granular activated carbon, PAC = powdered activated carbon, POTW = publicly owned treatment works, and RBC = rotating biological contactor; (—), wastewater; (- - -), return flows; (. . .), sludge.

A summary of available technologies is given in Tables 4 and 5. Attainable effluent qualities are shown in Table 6.

Table 4. Chemical Waste Treatment

Treatment method	Type of waste	Mode of operation	Degree of treatment	Remarks
ion exchange	plating, nuclear	continuous filtration with resin generation	demineralized water recovery; product recovery	may require neutralization and solids removal from spent regenerant
reduction and precipitation	plating, heavy metals	batch or continuous treatment	complete removal of chromium and heavy metals	one day's capacity for batch treatment; 3-h retention for continuous treatment; sludge disposal or de-watering required
coagulation	paperboard, refinery, rubber, paint, textile	batch or continuous treatment	complete removal of suspended and colloidal matter	flocculation and settling tank or sludge blanket unit; pH control required
adsorption	toxic or organics, refractory	granular columns of powdered carbon	complete removal of most organics	powdered carbon (PAC) used with activated sludge process
chemical oxidation	toxic and organics, refractory	batch or continuous ozone or catalyzed hydrogen peroxide	partial or complete oxidation	partial oxidation to render organics more bio-degradable

Table 5. Biological Waste Treatment

Treatment method	Mode of operation	Degree of treatment	Land requirements	Equipment	Remarks
lagoons	intermittent or continuous discharge; facultative or anaerobic	intermediate	earth dug; 10–60 days' retention (may require lining)		odor control frequently required groundwater considerations
aerated lagoons	completely mixed or facultative continuous basins	high in summer; less in winter	lined earth basin, 2.44–4.88 m deep; 8.55–17.1 m ³ (m ³ ·d)	pier-mounted or floating surface aerators or sub-surface diffusers	solids separation in lagoon; periodic de-watering and sludge removal
activated sludge	completely mixed or plug flow; sludge recycle	> 90% removal of organics	earth or concrete basin; 3.66–6.10 m deep; 0.561–2.62 m ³ (m ³ ·d)	diffused or mechanical aerators; clarifier for sludge separation and recycle	groundwater considerations excess sludge dewatered and disposed of
trickling filter	continuous application; may employ effluent recycle	intermediate or high, depending on loading	5.52–34.4 m ³ (10 ³ m ³ ·d)	plastic packing 6.10–12.19 m deep	pretreatment before POTW or activated sludge plant

RBC	multistage continuous	intermediate or high	$6.24 \times 10^{-7} \text{ } 4.68 \times 10^{-6} \text{ m}^{-1} \times \text{m}^2$	plastic disks	solids separation required
anaerobic	complete mix with re-cycle; upflow or down-flow filter, fluidized bed; upflow sludge blanket	intermediate		gas collection required; pretreat-ment before POTW or activated sludge plant	
spray irri-gation	intermittent application of waste	complete; water percolation into groundwater and runoff to stream		aluminum irrigation pipe and spray nozzles; movable for relocation	solids separation required; salt content in waste limited

Table 6. Maximum Quality Attainable From Waste Treatment Processes

Process	BOD	COD	SS	N	P	TDS
sedimentation, % removal	10–30		59–90			
flotation ^a , % removal	10–50		70–95			
activated sludge, mg/L	<25	^b	<20	^c	^c	
aerated lagoons, mg/L	<50		<50			
anaerobic ponds, mg/L	>100		<100			
deep-well disposal	^d					
carbon adsorption, mg/L	<2	<10	<1			
denitrification and nitrification, mg/L	<10			<5		
chemical precipitation, mg/L			<10		<1	
ion exchange, mg/L			<1	^e	^e	^e

^a Higher removals are attained when coagulating chemicals are used.

^b $\text{COD}_{\text{inf}} = [\text{BOD}_{\text{ult}}(\text{removed } 0.9)]$.

^c $\text{N}_{\text{inf}} = 0.12$ (excess biological sludge), kg; $\text{P}_{\text{inf}} = 0.026$ (excess biological sludge), kg.

^d Total disposal of waste.

^e Depends on resin used, molecular state, and efficiency desired.

Pre- or Primary Treatment

The principal objectives of pretreatment are to remove heavy metals prior to subsequent treatment, to neutralize the wastewater to a suitable pH for discharge or subsequent treatment; to remove high concentrations of suspended solids, to eliminate or reduce toxicity, and to eliminate or reduce volatiles. The concentrations of various pollutants that make pretreatment desirable are summarized in Table 7.

Table 7. Concentration of Pollutants That Make Prebiological Treatment Desirable

Pollutant or system condition	Limiting concentration	Kind of pretreatment
suspended solids, mg/Ld	>125	sedimentation, flotation, lagooning
oil or grease	>35	skimming tank or separator
toxic ions, mg/L		precipitation or ion exchange
Pb	<0.1	
Cu + Ni + CN	<1	
Cr ³⁺ + Zn	<3	
Cr ³⁺	<10	
pH	6 to 9	neutralization
alkalinity	^a	neutralization for excessive alkalinity
acidity	^b	neutralization
organic load variation	>2:1	equalization
sulfides, mg/L	>100	precipitation or stripping with recovery
ammonia, mg/L as N	>500	dilution, ion exchange, pH adjustment and stripping
temperature, °C in reactor	>38	cooling

^a Alkalinity of 0.5 kg as CaCO₃, BOD removed.

^b Free mineral acidity.

Equalization. The objective of equalization is to reduce variability in flow or strength of industrial wastewaters so that they can be successfully treated using either biological or physical–chemical treatment processes. It is axiomatic that an increase in influent concentration will result in an increase in effluent concentration. A doubling of influent concentration will not necessarily result in a doubling of effluent concentration. If the wastewater is readily degradable, increased biological activity will reduce the effluent concentration. If inhibitory constituents exist, however, a doubling of the influent concentration may more than double the effluent concentration. It is therefore not possible to predict final effluent quality without a predication of influent quality. If the wastewater flow is fairly constant, as in a pulp and paper mill, a constant-volume basin can be employed in which wastewater strength is equalized.

If the wastewater flow and strength are highly variable, as in a batch process chemical plant, a variable-volume basin with a variable inflow and a constant outflow is employed to equalize both flow and strength.

If the wastewater is readily degradable, as in a brewery, aeration is provided in the equalization basin to avoid septicity and the generation of odors.

In order to handle accidental spills or overflows, a spill basin may be provided, into which the flow is diverted if the concentration of a particular constituent exceeds a predetermined value. If equalization precedes biological treatment, in addition to the organic loading, high fluctuations in temperature, salinity, and toxic organics must also be considered. After the spill is contained, the wastewater flow is diverted back to the equalization basin. The contents of the spill basin are then pumped at a constant controlled rate to the equalization basin.

Neutralization. Wastewater discharge usually requires a pH between 6 and 9. Exceptions are a biological process in which microbial respiration degrades acidity (acetic acid is oxidized to CO₂ and H₂O), or one in which the CO₂ generated by microbial respiration neutralizes caustic alkalinity (OH⁻) to bicarbonate HCO₃⁻.

Neutralization usually follows equalization so that acidic and alkaline streams can be partially neutralized in the equalization basin. If the wastewater is always acidic, neutralization may occur at a stage prior to the stream reaching the equalization basin, so as to minimize corrosion in the equalization basin.

Acidic wastewaters can be neutralized with lime, magnesium hydroxide, caustic, or limestone. Lime or magnesium hydroxide is preferred over caustic because it is lower in cost and usually produces a more dewaterable sludge. A limestone bed is simple to operate and is applicable to moderately acidic wastewaters. The wastewater should have a near-constant acidity in order to maintain a constant effluent acidity. Highly acidic wastewaters require a two-stage process because of the logarithmic nature of pH. The first stage adjusts the pH to 3–3.5, and the second stage trims the pH to 6.5–7.5. A two-stage neutralization process is shown in Figure 2.

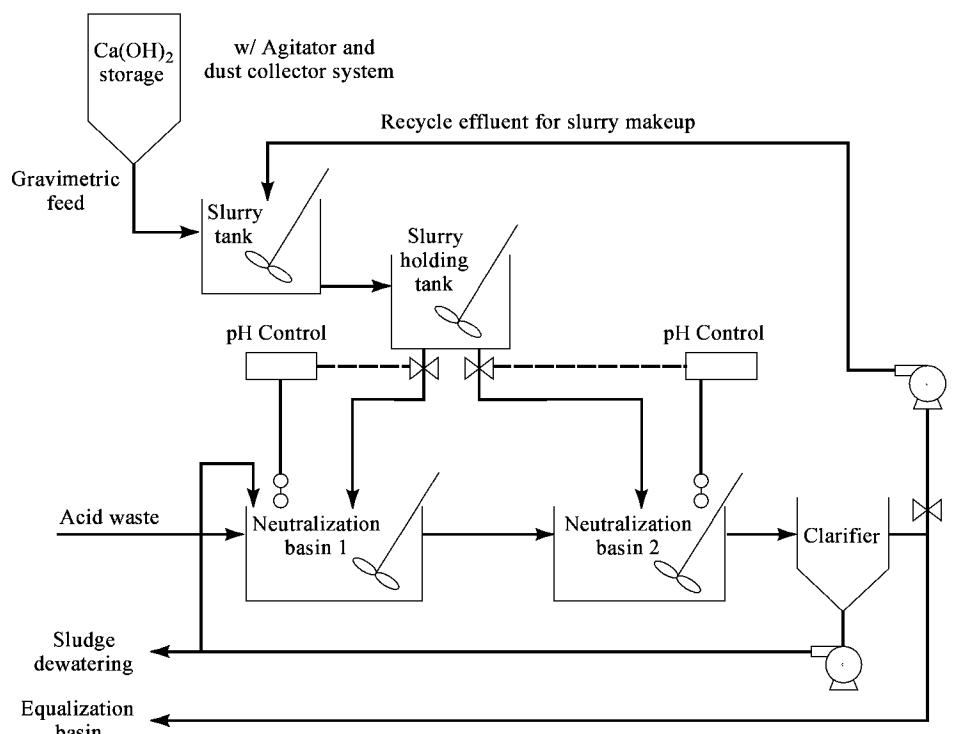


Fig. 2. Schematic diagram of a two-stage pH neutralization system.

Alkaline wastewaters can be neutralized with H₂SO₄ or HCl, or by using flue gas (CO₂).

Removal of Oil and Grease. High concentrations of oil and grease can be removed in a gravity separator where the lighter oils and greases float to the surface, where they are skimmed off. The API gravity separator removes oil globules of 0.015 cm or greater and can achieve an effluent oil content of less than 50–100 mg/L. The corrugated plate separator with a narrow separation space can remove oil globules of 0.01 cm or greater. As a result, effluent oil concentrations as low as 10 mg/L are achievable. A corrugated plate oil separator is shown in Figure 3.

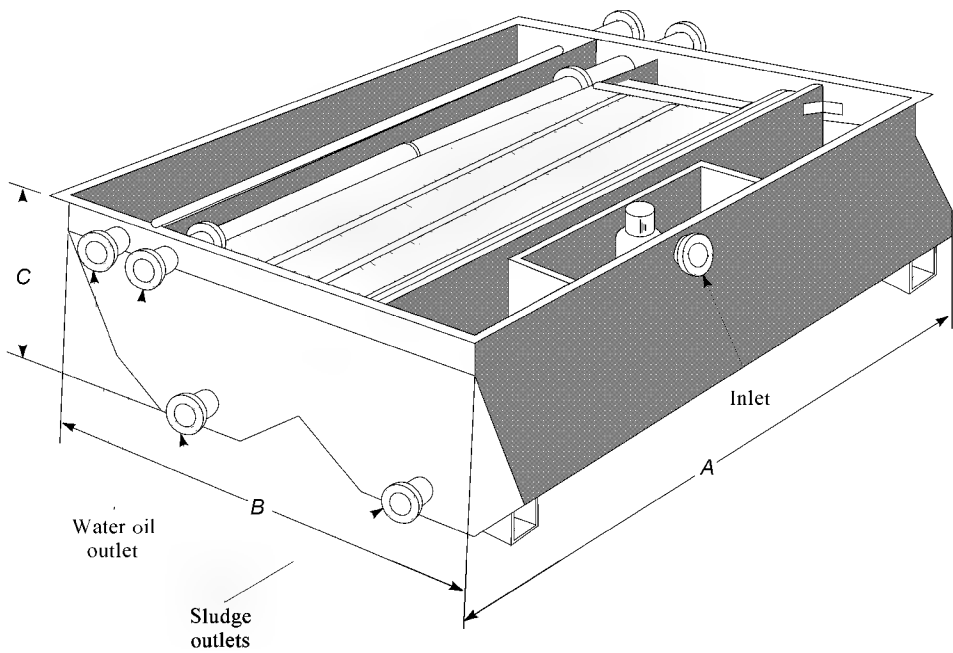


Fig. 3. Cutaway of a corrugated plate oil separator.

Low concentrations of oil can be removed by dissolved air flotation (DAF). In this process, an effluent recycle is pressurized in the presence of excess air, causing additional air to go into solution, in accordance with Henry's Law. When this water is discharged to the inlet chamber of the flotation unit at close to atmospheric pressure, the dissolved air comes out of solution in the form of tiny air bubbles which attach themselves to and become enmeshed in suspended solids and oil globules. The primary design criteria is the air solids ratio, which is defined as the mass of air released divided by the mass of solids fed. Sufficient air must be released to capture the solids in the influent wastewater. The performance of DAF for the treatment of several

wastewaters is shown in Table 8. In cases where the oil globules are of a very small size, a coagulant, usually alum, and a polymer are added to flocculate the particles, thereby enhancing bubble attachment and flotation. A DAF system is shown in Figure 4.

Table 8. Air Flotation Treatment of Oily Wastewaters

Wastewater	Coagulant, mg L	Oil concentration, mg L		
		Influent	Effluent	Removal, %
refinery	0	125	35	72
	100 alum	100	10	90
	130 alum	580	68	88
	0	170	52	70
oil tanker ballast water	100 alum + 1 mg L polymer	133	15	89
paint manu-facture	150 alum + 1 mg L polymer	1,900	0	100
aircraft mainte-nance	30 alum + 10 mg L activated silica	250–700	20–50	90 +
meat packing		3,830	270	93
		4,360	170	96

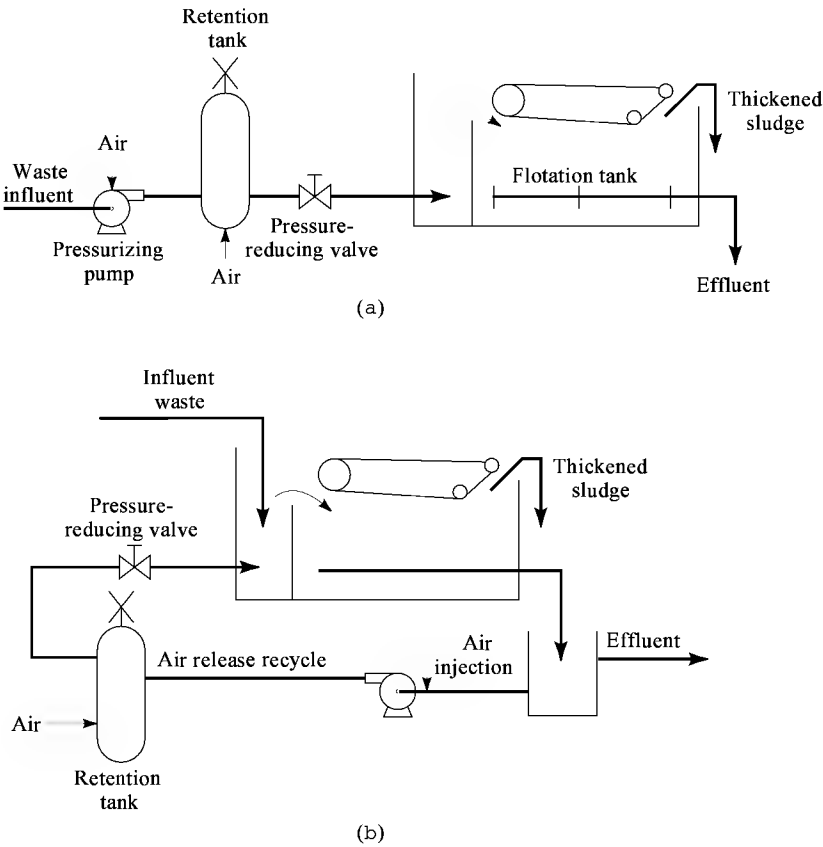


Fig. 4. Schematic of a pressurized dissolved air flotation system. (a) Influent pressurization; (b) recycle pressurization.

Alternatively, induced air flotation (IAF) can be employed, in which air bubbles are generated through an inductor. The removal mechanism is the same as the DAF.

Emulsified oil contains a liquid film so that it will not separate by gravity without first breaking the emulsion. This is achieved by adding surfactants, emulsion breaking polymers or coagulants. After the emulsion is broken, the conventional technologies described above are applicable.

Suspended Solids Removal. Depending on the concentration and characteristics of the suspended solids, they can be removed by filtration, flotation, or sedimentation. Coarse solids are removed by screening. Settleable suspended solids are removed in a clarifier, which may be circular or rectangular. The efficiency of solids removal is a function of the overflow rate ($\text{m}^3 \text{ m}^{-2} \cdot \text{d}$ ($\text{gal ft}^2 \cdot \text{d}$) as shown in Figure 5.

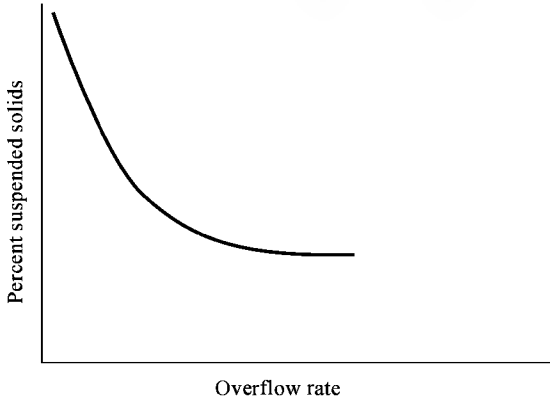


Fig. 5. Relationship between suspended solids removal design parameters and overflow rate.

Inorganic and organic colloidal suspensions in wastewater can be removed by chemical coagulation. Coagulation has been defined as the addition of

a positively charged ion such as Al^{3+} , Fe^{3+} , or cationic polyelectrolyte that results in particle destabilization and charge neutralization. Coagulation involves the formation of complex oxides that form flocculent suspensions which subsequently are separated from the liquid by sedimentation. Examples include colloidal dispersions of turbidity and color. The coagulants commonly employed are alum (Al_2SO_4), iron (FeCl_3), and sometimes lime (CaOH_2). Cationic or anionic polyelectrolytes are also used to enhance coagulation. A sludge blanket unit with sludge recirculation that will frequently result in lower coagulant requirements and enhanced clarification is shown in Figure 6.

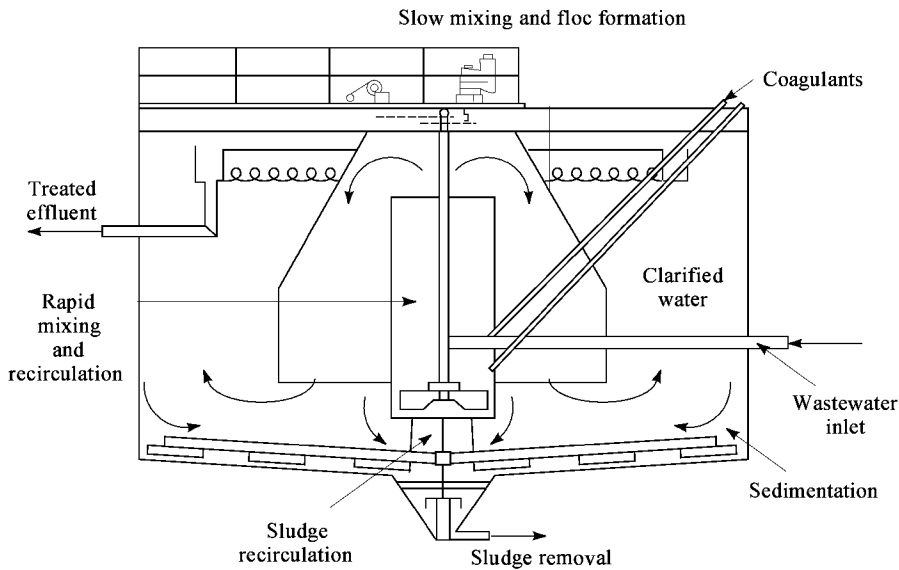


Fig. 6. A reactor clarifier designed for both coagulation and settling.

The application of coagulation for removal of dispersed and colloidal solids of several industrial wastewaters is shown in Table 9.

Table 9. Coagulation of Industrial Wastewaters

Wastewater	Parameter	Influent, mg L	Effluent, mg L
board mill	BOD	127	68
	TSS	593	44
tissue mill	BOD	140	36
	TSS	720	15
ball bearing	TSS	544	40
	O G	302	28
laundry	ABS	63	0.1
	BOD	243	90
latex	COD	4,340	178
	BOD	1,070	90
	total solids	2,550	446

Filtration is employed when the suspended solids concentration is less than 100 mg L and high effluent clarity is required. Finely dispersed suspended solids require the addition of a coagulant prior to filtration. Filters most commonly used in wastewater treatment are a dual media (anthraflit and sand) or a moving bed or continuous-backwash sand filter. Performance data for the tertiary filtration of municipal and industrial wastewater are shown in Table 10.

Table 10. Filtration Performance

Filter type	Wastewater	Filter depth, m	Hydraulic loading, L (min·m ²)	Removal, %		Effluent, mg L	
				SS	BO D	SS	BO D
gravity downflow	TF effluent	0.61–0.91	122	67	58		2.5
pressure upflow	AS effluent	1.52	90	50	62	7.0	6.4
dual media	AS effluent	0.76	204	74	88	4.6	2.5
gravity downflow	AS effluent	0.30	216	62	78	5	4
dynasand	metal finishing	1.01	163–244	90		2–5	
	AS effluent	1.01	122–407	75–90		5–10	
	oily wastewater	1.01	81–244	80–90 ^a		5–10 ^a	
hydroclear	poultry	0.30	81–204	88		19	
	oil refinery	0.30	81–204	68		11	
	unbleached kraft	0.30	81–204	74		17	

^a Free oil.

Heavy Metals Removal. Heavy metals should be removed prior to biological treatment or use of other technologies which generate sludges to avoid comingling metal sludges with other, nonhazardous sludges.

Technologies Available. The technologies available for metals removal are summarized in Table 11. In the case of precipitation of the metals as hydroxide, lime or caustic is added, usually to the pH of minimum solubility. This process has limitations in cases where multiple metals are present with a pH range of minimum solubility. Another technology is precipitation as the carbonate. Precipitation as the sulfide has the advantage of having a wide range of minimum solubility. A disadvantage is that a poorer thickening sludge is generated and sulfide presents a potential odor and health hazard. Solubility relationships are shown in Figure 7.

Table 11. Heavy Metals Removal Technologies

Process type	Examples
conventional precipitation	hydroxide sulfide carbonate
enhanced precipitation	coprecipitation dimethyl thio carbamate diethyl thio carbamate trimercapto- <i>s</i> -triazine, trisodium salt
other methods	ion exchange adsorption
recovery opportunities	ion exchange membranes electrolytic techniques

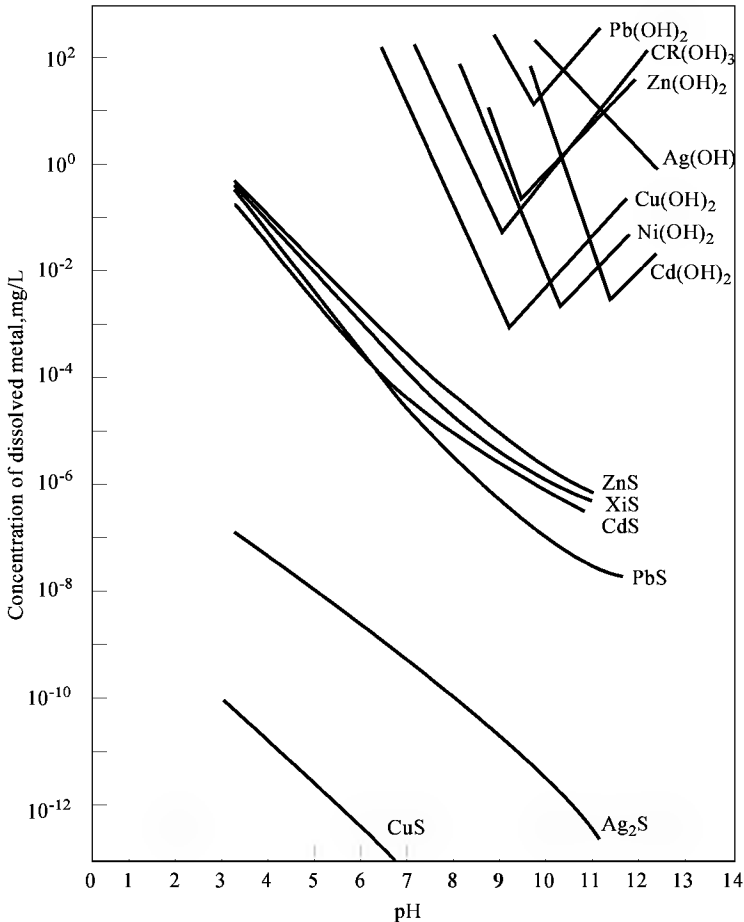


Fig. 7. Solubility of metal hydroxides and sulfides as a function of pH.

As a result, sulfide precipitation is frequently used as a polishing step following hydroxide precipitation.

Coprecipitation is a technology by which many metals such as arsenic will adsorb on alum or iron flocs and be effectively removed over a near-neutral pH range. The disadvantage of coprecipitation is the generation of large quantities of sludge.

New chelating ion-exchange resins are able to selectively remove many heavy metals in the presence of high concentrations of univalent and divalent cations such as sodium and calcium. The heavy metals are held as weakly acidic chelating complexes. The order of selectivity is Cu > Ni > Zn > Co > Cd > Fe²⁺ > Mn > Ca. This process is suitable for end-of-pipe polishing and for metal concentration and recovery.

Many heavy metals are removed on activated carbon. A primary mechanism is sulfide precipitation on the carbon.

Reverse osmosis (RO) can be employed to remove and recover heavy metals, particularly nickel.

Achievable effluent concentrations of heavy metals is summarized in Table 12.

Table 12. Effluent Levels Achievable in Heavy Metal Removals

Metal	Achievable effluent concentration, mg L	Technology
arsenic	0.05 0.06 0.005	sulfide ppt with filtration carbon adsorption ferric hydroxide co-ppt
barium	0.5	sulfate ppt
cadmium	0.05 0.008	hydroxide ppt at pH 10–11 co-ppt with ferric hydroxide sulfide precipitation
copper	0.02–0.07 0.01–0.02	hydroxide ppt sulfide ppt
mercury	0.01–0.02	sulfide ppt

	0.001–0.01	alum co-ppt
	0.0005–0.005	ferric hydroxide co-ppt
	0.001–0.005	ion exchange
nickel	0.12	hydroxide ppt at pH 10
selenium	0.05	sulfide ppt
zinc	0.1	hydroxide ppt at pH 11

Removal of Volatile Organics. Volatile organics, eg, benzene and toluene, should usually be removed prior to biological treatment.

The removal of volatile organics by air stripping is accomplished in packed or tray towers in which air is introduced to the bottom of the tower counterflow to the liquid passing down the tower. Stripping can also be accomplished using diffused or surface aeration. Removal of volatiles is a function of Henry's constant, the air liquid ratio, and the transfer efficiency of packing. High concentrations of volatiles require treatment of the off-gas through vapor phase carbon or combustion or a biofilter.

Volatile removal is a function of the air liquid ratio, and media height is shown in Figure 8. Typical stripping towers are shown in Figure 9.

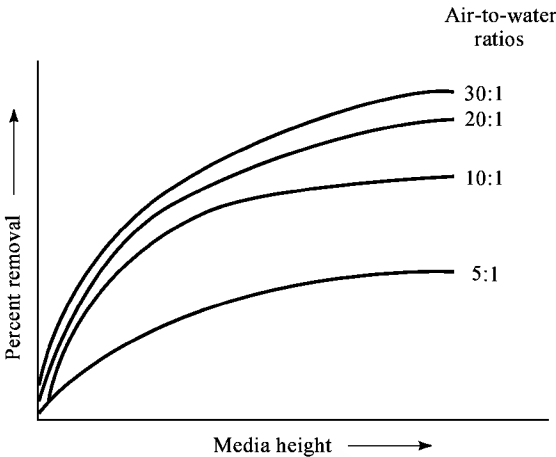


Fig. 8. Illustrative relationships of air stripping removal efficiencies to media height and air-to-water ratios.

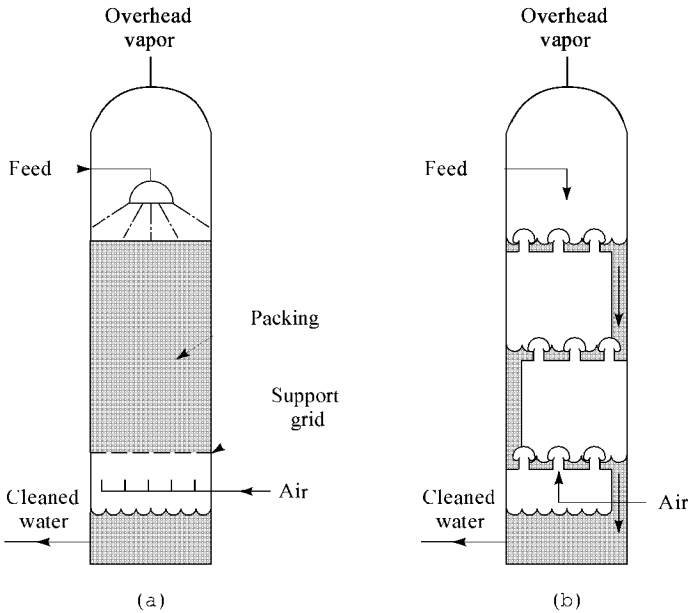


Fig. 9. Air stripping towers. (a) Packed tower; (b) tray tower.

In a steam stripper, steam is introduced into a packed tower, which causes volatiles to be removed in the vapor phase. An azeotropic mixture is formed, resulting in a separation of the volatiles from the water. An effluent recycle is usually employed to reduce volatiles in the liquid effluent.

Most volatile organics adsorb on activated carbon in the liquid state. Low concentrations of biodegradable volatiles can be removed by adsorption and biodegradation on activated carbon.

Many volatiles can be chemically oxidized using conventional or advanced chemical oxidants.

Biological Treatment

Aerobic treatment is generally applied to lower strength wastewaters, whereas anaerobic treatment is employed as a pretreatment for high strength wastewaters. The choice of process depends both on the concentration of organics and the volume of wastewater to be treated.

The objective of biological treatment is to remove biodegradable organics. In an aerobic biological treatment process, organic removal can occur through biodegradation, stripping, or sorption on the biological flocs.

Stripping. Degradable VOC, ie, benzene will both biodegrade and strip from the solution. The percentage stripped will depend on the power level in the aeration basin or the type of aeration equipment, ie, enhanced stripping with surface aerators. The biodegradation rate will decrease with increased halogenation, and hence the percentage stripped will increase.

Sorption. Most organics are sorbed to a very small degree on the biofloc, ie, < 2 percent. Exceptions are the nondegradable pesticide Lindane, other pesticides, and PCBs. Heavy metals will complex with the cell wall and precipitate within the floc. Metal accumulation will increase with increasing sludge age.

Biodegradation. The biodegradation properties of various organics are shown in Table 13. The mechanism of aerobic degradation is shown in

Figure 10.

Table 13. Relative Biodegradability of Certain Organic Compounds

Biodegradable organic compounds ^a	Compounds generally resistant to biological degradation
acrylic acid	ethers
aliphatic acids	ethylene chlorohydrin
	isoprene
aliphatic alcohols (normal, iso, secondary)	methyl vinyl ketone
aliphatic aldehydes	morpholine
aliphatic esters	oil
	polymeric compounds
alkyl benzene sulfonates with exception of propylene-based	polypropylene benzene sulfonates
benzaldehyde	
aromatic amines	selected hydrocarbons
dichlorophenols	aliphatics
ethanolamines	aromatics
glycols	alkyl–aryl groups
ketones	tertiary aliphatic alcohols
methacrylic acid	tertiary benzene sulfonates
methyl methacrylate	trichlorophenols
monochlorophenols	
nitriles	
phenols	
primary aliphatic amines	
styrene	
vinyl acetate	

^a Some compounds can be degraded biologically only after extended periods of seed acclimation.

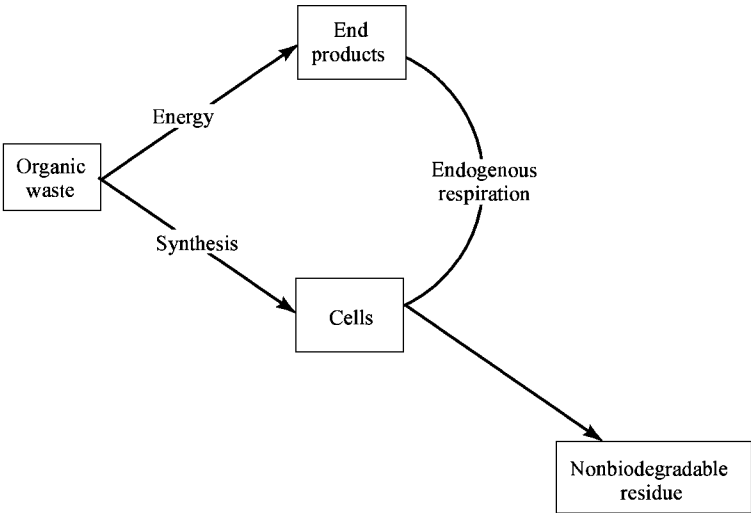
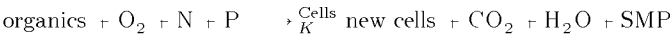


Fig. 10. The mechanism of aerobic biological oxidation.

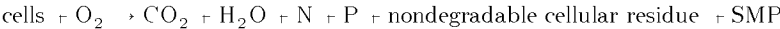
Approximately one-half of the organics removed are oxidized to CO₂ and H₂O, and one-half synthesized to biomass. Three to 10 percent of the organics removed result in soluble microbial products (SMP). The SMP is significant because it causes aquatic toxicity.

Nitrogen and phosphorus are required in the reaction at an approximate ratio of BOD:N:P of 100:5:1. Nitrogen and phosphorus are amply available in municipal wastewaters, but frequently are deficient in industrial wastewaters. It should be noted that only ammonia nitrogen or nitrate is available for biosynthesis.

The reactions in an aerobic biological process are as follows:



in which *K* is a reaction rate coefficient which is a function of the degradability of the wastewater and SMP is the nondegradable soluble microbial products and



The generation of SMP is directly proportional to the degradable COD removed in the process. Some of the SMP is toxic to aquatic species.

In the activated sludge process, performance is related to the food-to-microorganism ratio (F/M), which is the kg BOD applied/d/kg volatile suspended solids (VSS).

For a soluble wastewater, the VSS is proportional to the biomass concentration. Process performance may also be related to the sludge age, which is the average length of time the organisms are in the process.

$$\text{sludge age} = \frac{\text{mass of organisms under aeration}}{\text{mass wasted/d}}$$

Figure 11 depicts a batch oxidation. Note that readily degradable organics will be sorbed by the floc forming organisms immediately on contact. As organics are removed, oxygen is consumed and biomass is synthesized, as shown. Continued aeration after organic removal will result in oxidation of the biomass, generally referred to as endogenous respiration.

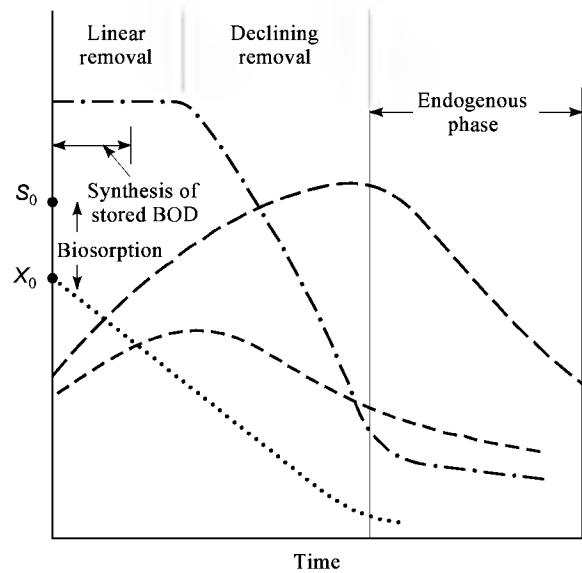


Fig. 11. Aerobic biological treatment, where S_0 = initial organic concentration and X_0 = initial biomass concentration; (—) total cell weight; (— · —) specific oxygen uptake rate; (---) cell N and P; (····) organic substrate remaining.

The performance, therefore, is related to the F/M or sludge age and the degradability, K . As the F/M decreases or the sludge age increases, greater removals are achieved. It should be noted that the sludge age is proportional to the reciprocal of the F/M . The reaction rate coefficient, K , as related to wastewaters characteristics.

As shown in Figure 10, all of the organics removed in the process are either oxidized to CO_2 and H_2O or synthesized to biomass generally expressed as volatile suspended solids. As previously noted, a small portion of the organics removed results in SMP products.

The fraction of the organics removed that result in synthesis varies, depending on nature and biodegradability of the organics in question. A rough estimate is to assume that one-half is oxidized and one-half synthesized.

For a soluble wastewater, the net sludge to be wasted from the process may be computed as synthesis minus oxidation, ie, endogenous respiration.

$$\text{net sludge wasted} = \text{sludge synthesized} - \text{oxidation}$$

As the sludge age is increased, more of the sludge is oxidized and the net sludge wasted is decreased. If the wastewater contains influent volatile suspended solids, such as that in a pulp and paper mill, the solids not oxidized in the process must be added to the net wasted.

The oxygen requirements are computed in a similar manner:

$$\text{oxygen required} = \text{organic oxidation} + \text{endogenous oxidation}.$$

Onaverage, it takes 1.4 kg oxygen to oxidize 1 kg cells as VSS. Therefore, for each kilogram of VSS subtracted from the sludge yield, 1.4 kg oxygen must be added to the oxygen required.

Process performance is affected by temperature. The reaction rate decreases with temperature over a range of 4–31°C. As the temperature decreases, dispersed effluent suspended solids increase. In one chemical plant in West Virginia, the average effluent suspended solids was 42 mg/L during the summer and 105 mg/L during the winter. Temperatures above 37°C may result in a dispersed floc and poor settling sludge. It is therefore necessary to maintain aeration basin temperature below 37°C to achieve optimal effluent quality.

Biological sludges generally fall into one of three classifications. A flocculent sludge is one in which the major part of the biomass consists of flocculent organisms, with some filaments growing within the floc. The advantage this provides is that the filaments form a backbone which strengthens the floc. Filamentous bulking occurs when the filaments grow out from the floc in the bulk of the liquid. This condition hinders sludge settling. The pinpoint case occurs at very low loadings, causing floc dispersion, as shown in Figure 12.

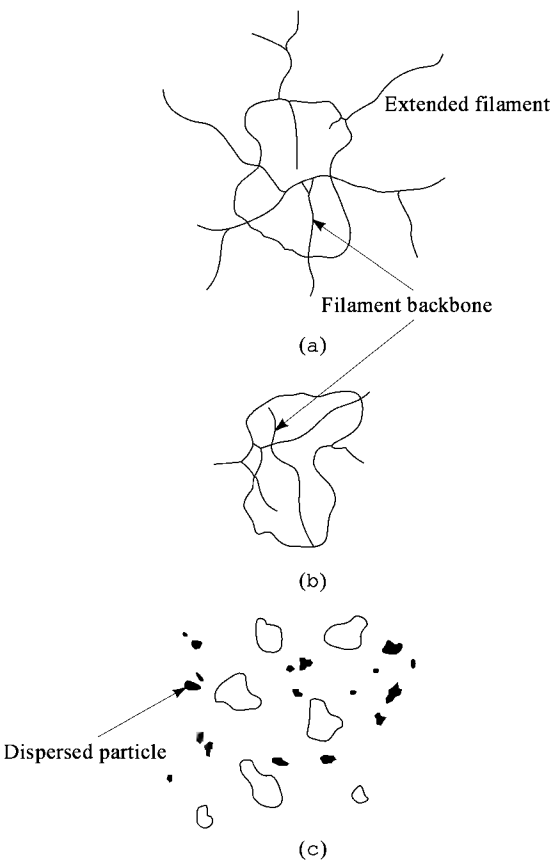


Fig. 12. Activated sludge types: (a) filamentous bulking; (b) nonbulking; (c) pinpoint.

Sludge quality is defined by the sludge volume index (SVI), ie, the volume occupied by one dry weight gram after settling for one-half hour, which therefore defines the bulkiness of the sludge. A bulking sludge is usually caused by an excess of filamentous-type organisms. Filamentous organisms thrive best with readily degradable organics as a food source. Wastewater containing complex organics is not subject to filamentous bulking because the filaments cannot degrade these organics. If all things are maintained equal, ie, adequate O₂, N and P, and BOD, the floc-forming organisms are predominant. In order to maintain conditions favorable to the floc formers, adequate oxygen, nutrients, and BOD must diffuse through the floc and reach all the organisms.

As the oxygen uptake or F/M increases, the dissolved oxygen must be increased to provide sufficient driving force to penetrate the floc. Minimum concentrations of nitrogen and phosphorus are necessary in the effluent.

Generalized flow configurations are shown in Figure 13. Refractory wastewaters can be treated in a complete mix basin because filamentous bulking is not an issue. For readily degradable wastewaters, high concentrations of BOD are necessary to penetrate the floc, requiring a plug flow configuration. Alternatively, a selector can be employed to absorb the readily degradable organics so they are not available as a food source for the filaments. The removal of specific priority pollutants follows the Monod kinetic relationship, which states that effluent quality is a function of sludge age. Thus the only way to reduce the effluent concentration of a specific organic is to increase the sludge age.

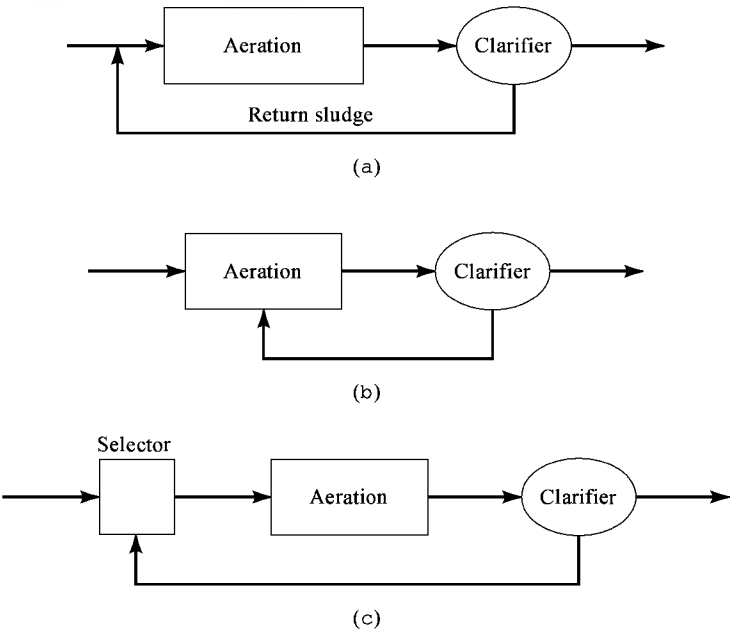


Fig. 13. Types of activated sludge processes: (a) plug flow; (b) complete mix; (c) selector-activated sludge.

Nutrient Removal

In many locations, nitrogen and phosphorus must be removed in order to meet effluent limitations.

Nitrogen. Nitrogen is most commonly removed through the process of biological nitrification and denitrification, as shown in Figure 14. In this process, organic nitrogen is hydrolyzed to ammonia. Ammonia, in turn, is oxidized to nitrite through the action of a specific organism, *Nitrosomonas*. This reaction also generates two hydrogen ions for each nitrogen oxidized so that alkalinity must be available to neutralize the acidity.

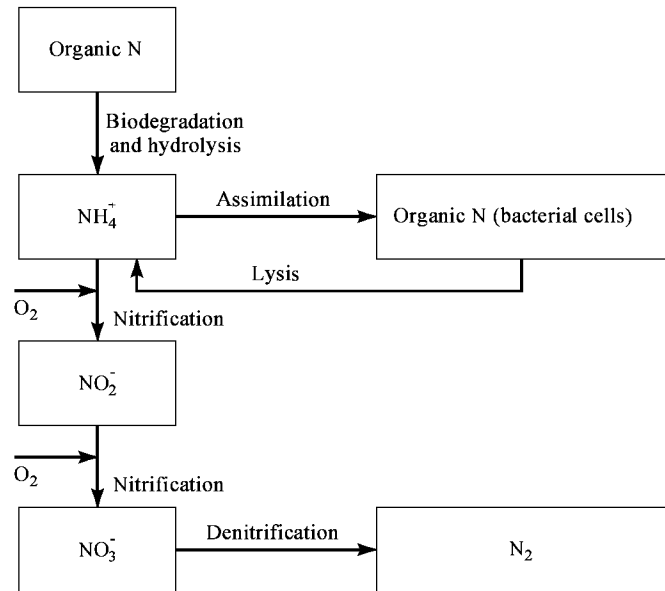


Fig. 14. Nitrogen transformation.

The nitrite, in turn, is oxidized to nitrate through the action of a specific organism, *Nitrobacter*. Since the *Nitrobacter* has a much higher growth rate than the *Nitrosomonas*, the critical design parameter is that the sludge age in the process must exceed the growth rate of the *Nitrosomonas*, or they will be washed out of the system.

As in all biological reactions, nitrification is a function of temperature. For municipal wastewaters, a minimum sludge age of 3.5 d is required at 20°C while the minimum sludge age must be increased to 12 d at 10°C in order to achieve nitrification.

The nitrifying organisms are relatively sensitive to many toxic organics, so that the treatment of industrial wastewaters requires special attention to the presence of toxics.

In one chemical plant, the required sludge age was found to be 22 d at 22°C, whereas for a nontoxic municipal wastewater, the required sludge age is 3 d.

Denitrification is a process in which facultative organisms will reduce nitrate to nitrogen gas in the absence of molecular oxygen. This consequently results in the removal of BOD₅. The denitrification process also generates one hydroxyl ion so that alkalinity requirements are reduced to half when both nitrification and denitrification are practiced.

The process of nitrification–denitrification can be practiced in one of two ways. In the oxidation ditch, nitrification occurs in the vicinity of the aerators. When the dissolved oxygen is depleted as the sludge–wastewater mixture passes from the aerator, denitrification occurs.

In the two-stage process, nitrification occurs under aerobic conditions in the second stage. The nitrified mixed liquor from the second stage is internally recycled to the anoxic first stage, where denitrification occurs.

High concentrations of ammonia can be removed by air or steam stripping. Ammonia can only be stripped in the nonionic form, NH₃. Since the amount of ammonia in the NH₃ form is a function of pH, a high (>10.5) pH is necessary for effective stripping. Ammonia can be oxidized by breakpoint chlorination to nitrogen gas. One disadvantage of this technology is the possible generation of chlorinated organic compounds other than ammonia and the production of chlorides, 10 parts of chloride for each ammonia oxidized. An ion-exchange resin specific for ammonia, clinoptilite, will remove ammonia.

Phosphorus Removal. Phosphorus can be removed from wastewater either chemically or biologically.
Chemical Removal. Phosphorus can be precipitated with lime to form Ca₃(PO₄)₂. The actual composition of the precipitate is a complex compound called apatite. Achieving minimum phosphorus concentrations requires a pH in excess of 10.5. Alum or iron will precipitate phosphorus as AlPO₄ or FePO₄. This procedure is generally employed in conjunction with the activated sludge process, in which the coagulant is added at the end of the aeration basin or between the aeration basin and the final clarifier.

Biological Removal. Certain organisms normally present in activated sludge have the ability to store phosphorus. The process configuration for bio-P removal involves an anaerobic step in which phosphorus is released and acetate taken up by the bio-P organisms. This is followed by an aerobic step in which phosphorus is rapidly taken up by the bio-P. Under proper operating conditions, soluble effluent phosphorus levels of 0.1 mg/L are achievable from municipal wastewater.

Alternative Biological Treatment Technologies

Lagoons. Where large land areas are available, lagooning provides a simple and economical treatment for nontoxic or nonhazardous wastewaters. There are several lagoon alternatives.
The impounding and absorption lagoon has no overflow or there may be an intermittent discharge during periods of high stream flow. These lagoons are particularly suitable to short seasonal operations in arid regions.

Anaerobic ponds are loaded such that anaerobic conditions prevail throughout the liquid volume. One of the major problems with anaerobic ponds is the generation of odors. The odor problem can frequently be eliminated by the addition of sodium nitrate at a dosage equal to 20% of the applied oxygen demand. An alternative is the use of a stratified facultative lagoon, in which aerators are suspended 3 meters below the liquid surface in order to maintain aerobic surface conditions, with anaerobic digestion occurring at the lower depths.

Aerobic lagoons depend on algae to produce oxygen by photosynthesis. This oxygen, in turn, is used by the bacteria to oxidize the organics in the wastewater. Since algae are aerobic organisms, the organic loading to the lagoons must be sufficiently low to maintain dissolved oxygen.

Aerated Lagoons. An aerated lagoon system is a two- or three-basin system designed to remove degradable organics (BOD). The first basin is fully mixed, thereby maintaining all solids in suspension. This maximizes the organic removal rate. A second basin operates at a lower power level, thereby permitting solids to deposit on the bottom. The solids undergo anaerobic degradation and stabilization. A third basin is frequently employed for further removal of suspended solids and enhanced clarification. In order to avoid groundwater pollution, these basins must usually be lined. The process is shown in Figure 15.

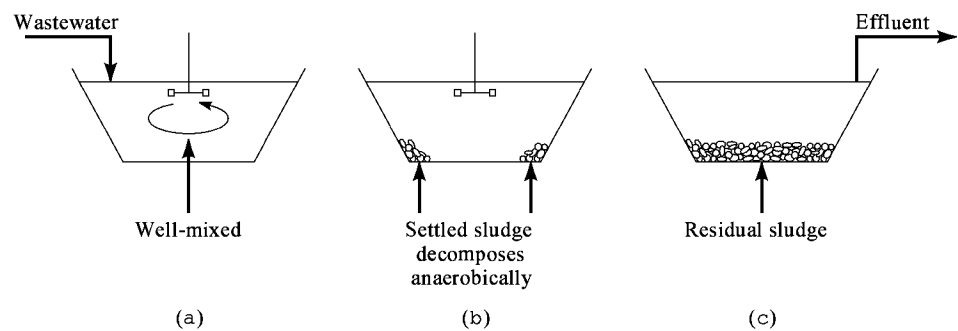


Fig. 15. Aerated lagoon types: (a) aerobic; (b) facultative; (c) settling.

Aerated lagoons are employed for the treatment of nontoxic or nonhazardous wastewaters such as food processing and pulp and paper. Retention time varies from 3 to 12 d, so a large land area is usually required.

Activated Sludge. There are several generic activated sludge processes presently available. Complete Mix (CMAS) is applicable to refractory-type wastewaters in which filamentous bulking is not a problem. This process has the advantage of dampening fluctuations of influent wastewater quality.

Plug flow is applicable for readily degradable wastewaters subject to filamentous bulking. Upstream controls are required to avoid shock loadings.

The selector process is applicable for readily degradable wastewaters; it also requires upstream controls. In a selector, degradable organics are removed by the floc formers by biosorption and therefore are not available as a food source for the filaments.

The sequencing batch reactor (SBR) or intermittent process is a combination of complete mix and plug flow, and usually controls filamentous bulking. The nature of the process eliminates the need for an external clarifier.

The oxidation ditch process is usually considered when nitrogen removal is required.

Other processes include deep tank aeration such as the Biohoch, the use of high purity oxygen, and the Deep Shaft process.

Performance of the activated sludge process may be summarized as follows: effluent quality is related to the sludge age, with higher sludge ages required for the more refractory wastewaters; degradable priority pollutants can be reduced to $\mu\text{g L}$ levels under optimal operating conditions as related to sludge age; effluent soluble BOD levels $<10\text{ mg L}$ are achievable in most cases, and nitrification and denitrification can be achieved through process modifications.

Fixed-Film Processes

Trickling Filter. A trickling filter is a packed bed, usually plastic on which a biofilm grows. As a wastewater passes over the film, organics and oxygen diffuse into the film, where they undergo biodegradation, as shown in Figure 16. The variables affecting performance are the organic loading rate, the hydraulic loading rate, temperature, and the degradability of the wastewater. For the treatment of industrial wastewaters, a trickling filter is considered a pretreatment process usually designed to remove about 50% of the BOD. This is largely the result of economic considerations. Trickling filter performance data are shown in Figure 17.

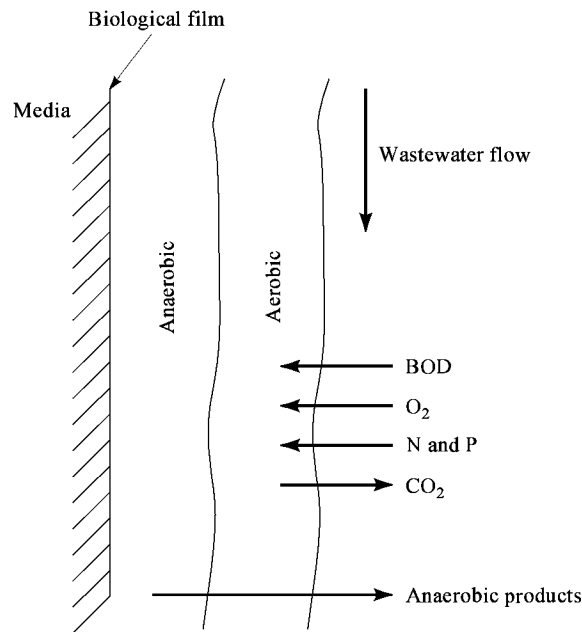


Fig. 16. Removal mechanisms in a fixed-film reactor.

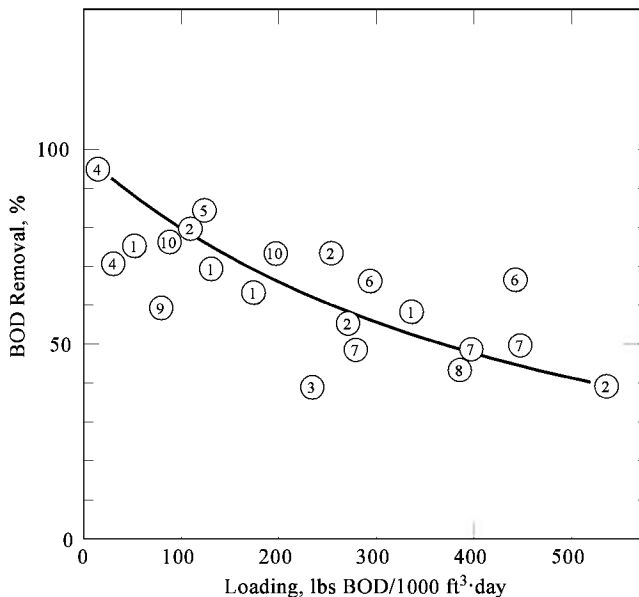


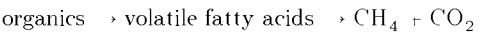
Fig. 17. Pretreatment of organic wastewater on trickling filters. Industry types for which coordinates have been plotted are 1, kraft pulp and paper; 2, mixed industry; 3, wet corn milling; 4, dairy; 5, tannery; 6, meat packing; 7, food; 8, pharmaceutical; 9, refinery; and 10, textile.

Rotating Biological Contactor (RBC)

An RBC is a fixed-film process in which a biofilm is developed on a rotating plastic cylinder which passes through the wastewater. As the cylinder passes through, the wastewater organics diffuse into the film. As the cylinder passes through the air, oxygen diffuses into the biofilm, causing degradation of the organics. Increased treatment is achieved by increasing the number of stages.

Anaerobic Treatment

Anaerobic treatment is usually employed for high strength wastewaters. In anaerobic treatment, complex organics are broken down through a sequence of reactions to end products of methane gas, CH₄, and carbon dioxide, CO₂:



Because anaerobic treatment will not reach usual permit discharge levels, it is employed as a pretreatment process prior to discharge to a POTW or to a subsequent aerobic process. Therefore it is most applicable to high strength wastewaters. Whereas aerobic treatment requires energy to transfer oxygen, anaerobic processes produce energy in the form of methane gas. Successful anaerobic process operation depends on maintaining a population of methane organisms. It is therefore critical that the sludge age of the anaerobic sludge exceed the growth rate of the methane organisms. At 35°C the common design criterion is a solids retention time (SRT) of 10 d or more. Anaerobic sludge can be maintained dormant for long periods of time, thereby making the process attractive for seasonal industrial operations such as in the food processing industry. A disadvantage to the anaerobic process is that initial startup may take as long as 45–60 d. Should the process be killed by a toxic shock a long period will be required for a re-startup. Particular care must be taken, therefore, to avoid upset. From an economic perspective, anaerobic pretreatment should be considered when the BOD exceeds 1000 mg L.

Types of Anaerobic Processes

There are five principal process variants which are propriety in nature. These are as follows:

- (1) Anaerobic Filter. The anaerobic filter is similar to a trickling filter in that a biofilm is generated on media. The bed is fully submerged and can be operated either upflow or downflow. For very high strength wastewaters, a recycle can be employed.
- (2) Anaerobic Contact. This process can be considered as an anaerobic activated sludge because sludge is recycled from a clarifier or separator to the reactor. Since the material leaving the reactor is a gas–liquid–solid mixture, a vacuum degasifier is required to separate the gas and avoid floating sludge in the clarifier.
- (3) Fluidized Bed. This reactor consists of a sand bed on which the biomass is grown. Since the sand particles are small, a very large biomass can be developed in a small volume of reactor. In order to fluidize the bed, a high recycle is required.
- (4) Upflow Anaerobic Sludge Blanket (UASB). Under proper conditions anaerobic sludge will develop as high density granules. These will form a sludge blanket in the reactor. The wastewater is passed upward through the blanket. Because of its density, a high concentration of biomass can be developed in the blanket.
- (5) ADI Process. The ADI is a low rate anaerobic process which is operated in a reactor resembling a covered football field. Because of the low rate, it is less susceptible to upset compared to the high rate processes. Its disadvantage is the large land area requirement.

With the exception of the ADI process, anaerobic processes usually operate at a temperature of 35°C. In order to maintain this temperature, the methane gas generated in the process is used to heat the reactor. Anaerobic processes are shown in Figure 18. Anaerobic treatment performance data are shown in Table 14.

Table 14. Performance of Anaerobic Processes

Wastewater	Process ^a	Loading, kg (m ³ ·d)	HRT, h	Temperature, °C	Removal, %
meat packing	anaerobic contact	3.2 (BOD)	12	30	95
meat packing		2.5 (BOD)	13.3	35	95
Keirring		0.085 (BOD)	62.4	30	59
slaughter house	upflow filter	3.5 (BOD)	12.7	35	95.7
citrus		3.4 (BOD)	32	34	87
synthetic		1.0 (COD)		25	90
pharmaceutical		3.5 (COD)	48	35	98

pharmaceutical		0.56 (COD)	36	35	80
guar gum		7.4 (COD)	24	37	60
rendering		2.0 (COD)	36	35	70
landfill leachate		7.0 (COD)		25	80
paper-mill foul condensate		10–15 (COD)	24	35	77
synthetic		0.8–4.0 (COD)	0.33–6	10–3	80
paper-mill foul condensate	expanded bed	35–48 (COD)	8.4	35	88
skim milk	UASB	71 (COD)	5.3	30	90
sauerkraut		8–9 (COD)			90
potato		24–45 (COD)	4	35	93
sugar		22.5 (COD)	6	30	94
champagne		15 (COD)	6.8	30	91
sugar beet		10 (COD)	4	35	80
brewery		95 (COD)			83
potato		10 (COD)			90
paper-mill foul condensate		4–5 (COD)	70	35	87
potato	ADI-BFV	0.2 (COD)	360	25	90
corn starch		0.45 (COD)	168	35	85
dairy		0.32 (COD)	240	30	85
confectionery		0.51 (COD)	336	37	85

^a UASB = upflow anaerobic sludge blanket.

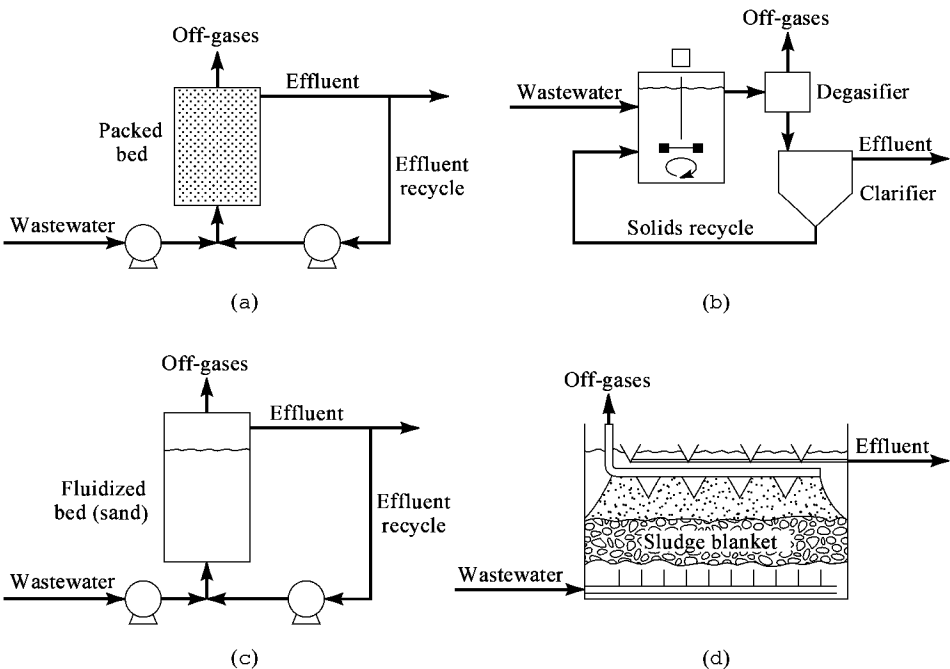


Fig. 18. Anaerobic wastewater treatment processes: (a) anaerobic filter reactor; (b) anaerobic contact reactor; (c) fluidized-bed reactor; (d) upflow anaerobic sludge blanket (UASB).

Advanced Wastewater Treatment

New regulations for toxics and priority pollutants frequently cannot be met by conventional technology. Other physical–chemical technologies must therefore be applied.

Chemical Oxidation. Chemical oxidation can be applied in industrial wastewater pretreatment for reduction of toxicity, to oxidize metal complexes to enhance heavy metals removal from wastewaters, or as a posttreatment for toxicity reduction or priority pollutant removal.

Complex organics and toxics are chemically oxidized to end products of CO₂ and H₂O or to intermediate products which are nontoxic and biodegradable.

The common oxidants are ozone, hydrogen peroxide, H₂O₂, catalyzed usually with ferrous iron, Fe²⁺, and in some cases chlorine dioxide and uv light. Advanced oxidation systems include H₂O₂ + uv; ozone + uv; and H₂O₂, ozone, and uv. Depending on the application, the oxidation can be complete to end products as in a contaminated groundwater or partial to degradable intermediate products as in a process wastewater.

Wet air oxidation is based on a liquid-phase oxidation between the organic material in the wastewater and oxygen supplied by compressed air. The reaction takes place flamelessly in an enclosed vessel which is pressurized and at a high temperature, typically 13.79 × 10³ Pa and 575°C. The system temperature, initiated by a start-up boiler, is maintained through auto thermal combustion of organics once the reaction starts.

Carbon Adsorption. Carbon can be employed either as granular carbon in columns (GAC) or as powdered carbon added to an activated sludge plant (PACT). Carbon removes most organics except low molecular weight soluble organics such as sugars and alcohols. In general, those organics which adsorb the poorest biodegrade the best, whereas those which biodegrade poorly adsorb well on carbon.

Design data are available for the specific organics on the EPA's priority pollutant list. For mixed wastewaters, a laboratory study is necessary to determine adsorption characteristics. Wastewater is contacted with a range of concentrations of powdered carbon and adsorption occurs, as graphed in the form of a Freundlich Isotherm, shown in Figure 19.

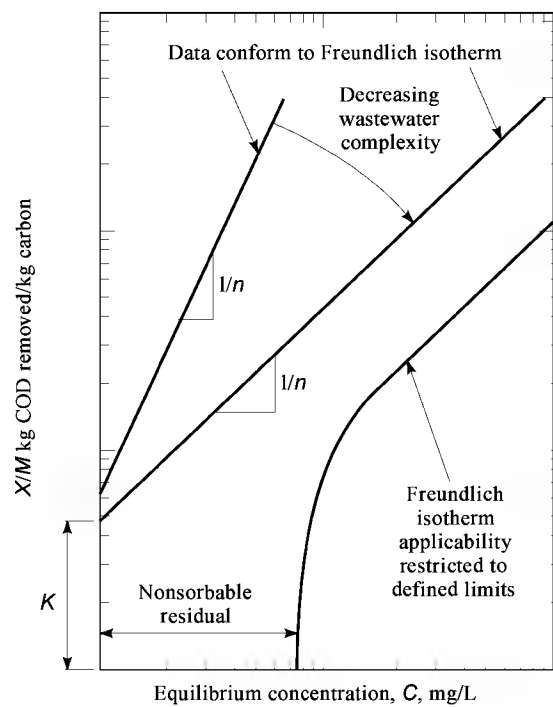


Fig. 19. Adsorption isotherm correlation, showing Freundlich isotherm application.

Because several commercial carbons are available, a comparative isotherm study should be made for a wastewater to determine which carbon is the best.

Granular carbon is regenerated. Regeneration can be accomplished by acid, caustic, solvent, steam, or thermal. For specific organics, it is usually feasible to employ one of the first four and operate *in situ*. For a typical multicomponent wastewater, however, thermal regeneration must be used. This can be accomplished using a multiple-hearth furnace or a fluidized-bed furnace. Attrition and oxidation losses range from 5 to 10 wt %. In addition, there is frequently a capacity loss, particularly for low molecular weight organics. Organic removal by activated carbon for several wastewaters is shown in Table 15.

Table 15. Granular Activated Carbon (GAC) Adsorption Effectively Removes Organics From a Variety of Industrial Wastewaters

Type of Industry	Wastewater TOC ^a or phenol, mg L, or color index		Average removal, %	Carbon usage nkg 10 ³ L
food	TOC	25–5,300	90	0.1–41
tobacco	TOC	1,030	97	6.9
textiles	TOC	9–4,670	93	0.12–29
	color	0.1–5.4	97	0.012–10
apparel	TOC	390–875	75	1.4–5.2
paper	TOC	100–3,500	90	0.4–19
	color	1.4	94	0.44
printing	TOC	34–170	98	0.52–0.55
chemicals	TOC	36–4,400	92	0.13–17
	phenol	0.1–5,325	99	0.20–22
	color	0.7–275	98	0.14–159
petroleum refining	TOC	36–4,000	92	0.13–17
	phenol	7–270	99	0.72–2.9
rubber and plastics	TOC	120–8,375	95	0.62–20
leather	TOC	115–9,000	95	0.36–38
stone, clay, glass	TOC	12–8,300	87	0.34–36
primary metals	TOC	11–23,000	90	0.06–222
fabricated metals	TOC	73,000	25	73

^a TOC = Total organic carbon.

Powdered activated carbon (PAC) is added at a stage prior to, or directly to, the activated sludge aeration basin, as shown in Figure 20. The sludge is therefore a mixture of biomass and carbon. The degradable organics are biodegraded, and the nondegradable organics are adsorbed on the carbon. PAC may also be applied to enhance nitrification.

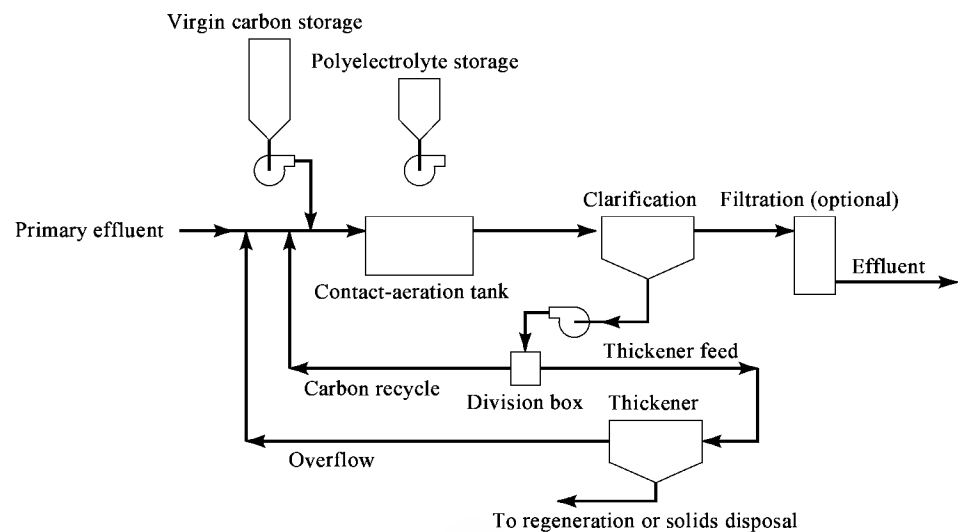


Fig. 20. Process flow diagram for PACT wastewater treatment system.

PAC sludge can be regenerated by wet air oxidation (WAO) or by a multiple-hearth furnace. Capacity losses might be high in WAO, particularly with low molecular weight organics. Weight loss in a furnace may exceed 20%.

Sludge Handling and Disposal

Types of Sludges. Municipal primary sludge consists of organic and inorganic particulates. The sludge must be stabilized before land disposal. Biological sludge consists of organisms and other particulates not degraded in the biological process. Chemical sludges consist of chemical precipitates, heavy metals, and other contaminants such as color precipitated from industrial wastewaters.

Sludge Stabilization. Organic sludges need to be stabilized before ultimate disposal except in the case of incineration. This is usually achieved by either aerobic or anaerobic digestion. In aerobic digestion, the degradable volatile solids are liquefied and oxidized to CO₂ and H₂O. In anaerobic digestion the solids are liquefied and fermented to CH₄ and CO₂.

Sludge Thickening. The mechanisms of sludge dewatering are shown in Figure 21. Thickening of sludge usually precedes dewatering. Depending on the nature of the sludge, several techniques are available for thickening.

- (1) Gravity thickening is applicable to primary municipal sludges and most chemical sludges.
- (2) Another technique is one in which the sludge is passed in a thin sheet over a porous drainage belt. This technique is particularly applicable to waste-activated sludge.
- (3) In dissolved air flotation, air bubbles float the sludge, which is then removed by a scraper. It is generally applicable to large volumes of waste-activated sludge.
- (4) In a centrifuge technique, various centrifuge types are used for sludge thickening.

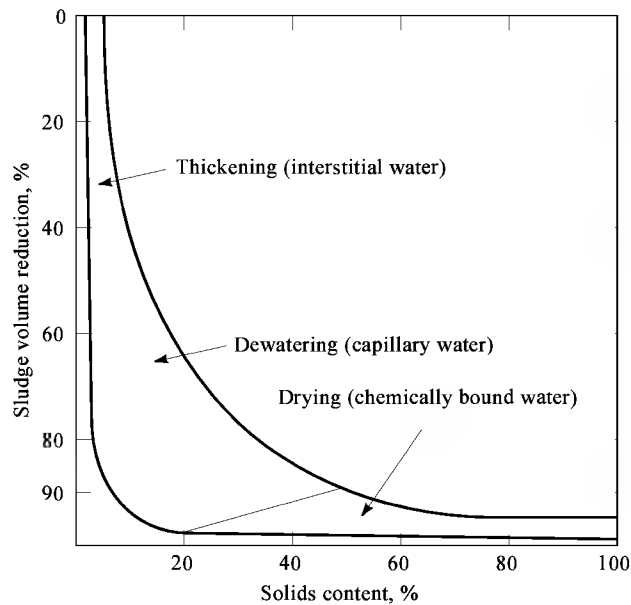


Fig. 21. Sludge thickening and dewatering relationships.

Sludge Dewatering. In a centrifuge technique for sludge dewatering, the solid bowl centrifuge concentrates the solids under centrifugal force. Both centrate and cake solids are continuously discharged from the machine. Polymer addition is required for most wastewater sludges.

A vacuum filter is a cloth-covered drum which operates under an applied vacuum. As the drum passes through the sludge vat, solids are deposited on the filter. As the drum passes through the air, drying of the cake occurs. The cake is continuously discharged to a conveyor belt.

A belt filter press consists of a gravity drainage belt, followed by a series of roller presses which squeeze out water.

A pressure filter is a plate-and-frame press which operates on an intermittent time cycle. Drier cakes are generally attainable from a filter press.

Sludge drying beds are usually used for smaller sludge volumes, which drain and dry rapidly. Their application is usually restricted to the more arid climates.

Thickening and dewatering of various sludges is shown in Table 16.

Table 16. Performance of Sludge Thickening and Dewatering Equipment

Equipment, and type of sludge ^a	Loading	Resultant solids content, wt %
gravity thickener		
municipal WAS, kg m ² ·d	20–122	1–3
inorganic sludge, kg m ² ·d	122–366	10–20
flotation thickener		
municipal WAS, kg m ² ·h	15–29	4–7
centrifuge, per unit		
paper-mill WAS, L min	227–379	11
citrus-processing WAS, L min	95	9–10
vacuum filter		
municipal WAS, kg m ² ·h	10–39	10–15
belt filter press, per unit of belt width or area		
chemical WAS, kg m·h	149–343	13–17
paper-mill WAS, m ³ m·h	3–6	12–19
paper-mill primary sludge, m ³ m·h	12–30	18–37
meat processing WAS, m ³ m ² ·h	3.6	17
tannery WAS, m ³ m ² ·h	2.1	23
pressure filter		
chemical WAS		20–30
chemical WAS	4-h cycle	28
citrus-processing WAS	2-h cycle	27
tannery WAS, m ³ m ² ·h	0.09	48

^a WAS = waste-activated sludge.

Sludge Disposal. Land disposal of wet sludges can be accomplished in a number of ways: Lagooning or the application of liquid sludge to land by truck or spray system, or by pipeline to a remote agricultural or lagoon site.

In lagoons sludge is stored and in the case of organic sludges anaerobically digested. Odor control is achieved either by chemical addition to the overlying water (Cl₂ or H₂O₂) to oxidize sulfides, or by installing aerators in the liquid layer to maintain aerobic conditions.

Biological sludges can be incorporated into the soil. An important consideration is the heavy metal content of the sludge, which will dictate the total number of years sludge can be applied. The available nitrogen content of the sludge will determine the maximum yearly application.

Dewatered sludges can be employed as a landfill.

Incineration can be accomplished in multiple-hearth furnaces, in which the sludge passes vertically through a series of hearths. In a fluidized-bed sludge, particles are fed into a bed of sand fluidized by upwardly moving air.

Storm-Water Control

In most industrial plants, it is now necessary to contain and control pollutional discharges from storm water. Pollutional discharges can be minimized by providing adequate diking around process areas, storage tanks, and liquid transfer points, with drainage into the process sewer. Contaminated storm water is usually collected on the basis of a frequency for the area in question, eg, a 10-year storm, in a holding basin. The collected water is then passed through the wastewater treatment plant at a controlled rate.

BIBLIOGRAPHY

"Wastes, Industrial" in *ECT* 1st ed., Vol. 14, pp. 896–914, by H. E. Orford, W. B. Snow, and W. A. Parsons, Rutgers University; in *ECT* 2nd ed., Vol. 21, pp. 625–651, by R. D. Ross, Thermal Research & Engineering Corp.; in *ECT* 3rd ed., Vol. 24, pp. 227–256, by N. L. Nemerow, University of Miami.

1. R. A. Conway and R. D. Ross, *Handbook of Industrial Waste Disposal*, Van Nostrand Reinhold Co., Inc., New York, 1980, p. 125.

W. Wesley Eckenfelder, Jr.
Eckenfelder Inc.

WATER

- Sources and quality issues,
- Properties,
- Pollution,
- Analysis,
- Supply and desalination,
- Industrial water treatment,
- Municipal water treatment,
- Sewage,
- Reuse,
- Treatment of swimming pools, spas, and hot tubs,
- Waterproofing and water oil repellency,

SOURCES AND QUALITY ISSUES

The physical effectiveness of water moving through the hydrologic cycle, as a means of modification of surface topography, is impressively demonstrated in such geographic erosional features as the Grand Canyon of the Colorado River. Besides the physical movement of erodible material, however, the chemical attack of circulating water on rock minerals over time brings into solution a wide variety of elements that are transported to the oceans by rivers.

In the early years of development of the science of geochemistry there was a strong interest by investigators Goldschmidt and Vernadski in Europe and Clarke in the United States in the significance of chemical composition of river water and how the observed composition was reached through water-rock chemical reaction. There also were direct practical motivations of studies of river water composition. For example, during the nineteenth century the settlement of the interior of the United States was continuous. The growing major cities commonly were located on navigable streams, or at the shoreline of other surface water bodies, that could be used both as sources of water supplies and for disposal of wastes as well as for transportation. That these uses were commonly not compatible became clear as populations of cities increased.

Evaluation of water for its fitness to be used as a public supply, the procedures generally termed "sanitary analysis," involve such determinations as dissolved oxygen, biochemical oxygen demand, testing for presence of various forms of bacteria, and related evaluations, are described in "Standard Methods" (1) and are discussed only briefly in this article. The evaluation of river water for possible industrial uses is generally concerned with determination of major dissolved constituents in filtered sample aliquots. Samples must be taken at intervals frequent enough to show the degree of variability of these concentrations over a range of flow rates.

Ideally the historical record of stream water quality would extend back to a time when human activities in the drainage basin had no significant effects. This "pristine" condition had probably already passed in most U.S. rivers before any organized water quality studies were made, as concern about apparent stream pollution was commonly a motivating factor in starting such studies (see WATER, POLLUTION).

Factors Affecting Stream Water Quality

General Principles of Stream Water Geochemistry. The composition of pristine stream water can range from nearly that of rainwater, a very dilute solution containing up to a few tens of milligrams per liter of dissolved material, to concentrations of as much as several thousand milligrams per liter, which mainly are found in semiarid regions where saline springs contribute chemical constituents derived from readily soluble rock strata. In general, however, the dissolved-solids concentrations of principal streams in humid and subhumid regions of the conterminous United States have a narrower range. Analyses showing maximum and minimum concentrations for a recent 3-year period for 29 stream-sampling sites in the United States recently were published (2). Seven of these sites had maximum total dissolved-solids concentrations greater than 1000 mg L; these high concentrations are attributable to natural saline inflows, although human activities probably intensified their effects.

Chemical analyses of stream water that have been published since the early years of this century generally include determinations for four positively charged ions (cations)—calcium (Ca²⁺), magnesium (Mg²⁺), sodium (Na⁺), and potassium (K⁺)—and five negatively charged ions (anions)—bicarbonate (HCO₃⁻), sulfate (SO₄²⁻), chloride (Cl⁻), fluoride (F⁻), and nitrate (NO₃⁻)—and uncharged dissolved silicic acid (generally reported in terms of silica, SiO₂). These are the major constituents in most natural stream water and are those given principal attention in this article. Minor constituents, present at concentrations substantially less than 1 mg L, include a wide variety of inorganic and organic constituents, and although these constituents commonly are the result of human activities, they are not considered in detail in this article because few reliable measurements were made until the 1970s. Short-term trends for these constituents have been evaluated using statistical techniques (3).

Chemical Reactions that Govern Stream Composition. Weathering is a general term for mechanical and chemical alteration of rock minerals that are exposed to the atmosphere and circulating water at and near the land surface. Chemical reactions that occur during weathering produce both water-soluble and nonwater-soluble products; those that are water soluble are transported from the reaction site in surface runoff or in moving groundwater. To a certain extent, at least, the concentrations of dissolved elements would be expected to reflect the relative abundance of the elements in the rocks exposed at the reaction site. Such a broad generalization has some validity for silicon (Si) and the four elements that form the major cations of most natural stream water. These five elements are among the eight most abundant elements, with oxygen being the most abundant, in igneous and sedimentary rocks of the Earth's outer crust. The other two of the most abundant elements—aluminum (Al) and iron (Fe)—form oxides or hydroxides of very low solubility during normal rock weathering and, therefore, generally are not present in large amounts in stream water. On the other hand, major anions in stream water display a more complex relation to rock composition. In the average stream water sample, the five most abundant anions represent the nonmetallic elements carbon, sulfur, chlorine, fluorine, and nitrogen. Also, oxygen is included in three of the anions of these elements.

Oxygen is by far the most abundant element in crustal rocks, composing 46.6% of the lithosphere (4). In rock mineral structures, the predominant anion is O²⁻, and water (H₂O) itself is almost 90% oxygen by weight. The nonmetallic elements fluorine, sulfur, carbon, nitrogen, chlorine, and phosphorus are present in lesser amounts in the lithosphere. These elements all play essential roles in life processes of plants and animals, and except for phosphorus and fluorine, they commonly occur in earth surface environments in gaseous form or as dissolved anions.

In a very broad general sense, then, the major cationic constituents of stream water tend to reflect the composition of associated rocks and the relative resistance of the rock minerals to weathering. The anions, which must be present in these water solutions in electrochemical balance with the cations, tend to reflect the influence of various chemical and biochemical processes that have broken down the rock minerals as well as the chemical, biochemical, and physical processes taking place in the aquatic and surrounding environments. The predominance of bicarbonate anions in most stream water is related to cycling of carbon dioxide (CO₂) from air and to biological processes in soil. Dissolution of carbon dioxide in water produces carbonic acid (H₂CO₃) that attacks rock minerals. Bicarbonate anions are formed in solutions participating in such reactions in amounts equivalent to the amount of cations that are released.

Sulfur and nitrogen participate in biologically mediated oxidation reactions producing hydrogen ions (H⁺) that become available for weathering of rock minerals. For example, pyrite (FeS₂) can be converted to dissolved ferrous iron and sulfate as a result of oxidation of sulfur by dissolved oxygen, and the hydrogen ion is a major by-product. Carbonate or sulfate in sedimentary rocks (eg, limestone and gypsum) can be taken directly into solution and can add substantially to the bicarbonate and sulfate contents in water that is in contact with such rocks.

Chlorine plays a less significant role in chemical weathering processes than do sulfur and carbon. Most geochemists believe that much, or most, of the chloride in stream water in coastal areas is derived from sea salt that is carried landward or deposited by rainfall. Farther inland, however, a major part of the chloride loads in streams is the result of human activities.

The final composition of stream water is the product of the weathering reactions and related processes outlined above. However, the chemical processes are influenced and controlled by an intricate combination of environmental factors that are characteristic for each drainage system. Therefore, the composition of the bedrock in an area and the residual material left at the surface as soil and subsoil exert a strong influence on the chemical composition of runoff from the area. The reactions of water with this material are the ultimate geological control and are the source of soluble weathering products.

Most igneous and metamorphic rocks are composed predominantly of aluminosilicate minerals, including feldspar such as albite (NaAlSi₃O₈) or anorthite (CaAl₂Si₂O₈) and crystalline forms of silica such as quartz (SiO₂). Various mixed metal-plus-silicon oxides such as olivine [(Mg,Fe)₂(SiO₄)] and

pyroxene [Mg₂(SiO₃)₂] can be major constituents in darker colored igneous rocks that are relatively low in total silicon.

Rocks that were deposited as sediment can consist of unaltered fragments of a precursor rock body and are represented by sandstone and conglomerate. The finer grained sedimentary rocks, such as shale or siltstone, also can contain some unaltered particles but also usually have high proportions of slightly soluble alteration products, such as clay minerals, formed during weathering of resistant silicate minerals. Such rocks are classified as hydrolyzates. Another class of sedimentary rock of major importance is the precipitates, such as limestone and dolomite, which are predominantly composed of calcium carbonate and calcium plus magnesium carbonate, respectively. Evaporites are sedimentary rocks produced by extensive evaporation of water from weathering solutions. Common examples are gypsum and anhydrite, primarily composed of calcium sulfate, or halite (rock salt), primarily composed of sodium chloride. Obviously the more readily soluble minerals of evaporite or precipitate rocks can dissolve rapidly when exposed to circulating water. Carbonates also can act as cementing material between the mineral grains of resistate and hydrolyzate rocks.

The extent to which minerals are attacked and dissolved from igneous and metamorphic terranes depends in large part on the availability of reaction sites on solid surfaces and the length of time the solution–solid contact is maintained. The effects of weathering are controlled by kinetic factors, such as the rates of the chemical reactions and the general rates of water and sediment movement, and biological factors that include the effects of biotic growth in the weathering zone. Also, the hydrologic properties of the drainage system (precipitation, evaporation, runoff, slope of the area), the relative permeability of rocks and soils, and the degree to which the surface drainage system is coupled to the groundwater reservoirs are important modulating forces. In rocks that contain more soluble minerals such as calcite, the degree to which solids are dissolved and carried off in the runoff is more likely to be governed by chemical thermodynamic factors, and in carbonate systems a state of chemical equilibrium could be closely approached.

Mineral dissolution reactions of importance generally require a continuous supply of hydrogen ions in the incoming solution. To some degree, reacting hydrogen ions are supplied from the water itself, which always includes, to some extent, water molecules that have broken apart (dissociated) into hydrogen and hydroxide ions. Under standard conditions (25°C and 1 atm pressure), the effective concentration (activity) of hydrogen ions in pure water is 10^{-7.00} mol/L. (A mole of a chemical element is a quantity in grams numerically equal to the atomic weight. For hydrogen, 10^{-7.00} moles/L is equivalent to 0.1 µg/L.)

The pH scale commonly used to express acidity is defined as the negative base-10 logarithm of the activity of hydrogen ions in a solution. At neutrality under the conditions defined in the preceding paragraph, the pH therefore will be 7.00. A change of 1 pH unit represents a 10-fold change in hydrogen ion activity. The activity of a dissolved ion is exactly equal to its concentration only in very dilute solutions. This topic has been discussed and explained more extensively elsewhere (5).

In natural systems the most effective sources of hydrogen ions generally are chemical reactions involving dissolved constituents. An important source is carbon dioxide gas, which is present in weathering solutions as a result of contact with air; it is produced in larger quantities by plant root respiration and decay of soil organic matter and by the metabolic processes of various organisms in water and sediment. Equation 1 shows that some of the carbon dioxide that dissolves forms carbonic acid (note that in the following equations the arrows indicate the direction in which the reaction normally proceeds; reactants are on the left side and products on the right; double arrows indicate that reactions can proceed in either direction):



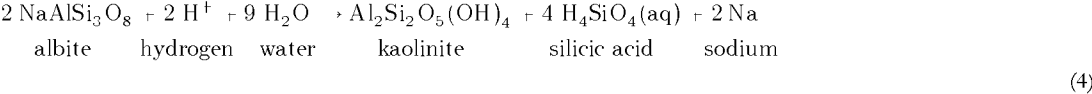
Equation 2 shows that the acid dissociates to form bicarbonate ions and hydrogen ions:



and carbonate anions can be formed in a second dissociation step:



The H⁺ that is supplied by reactions in Eqs. 2 and 3 can react with silicate minerals such as the sodium-bearing form of feldspar:



to produce the clay mineral kaolinite, undissociated silicic acid (H₄SiO₄ which also can be written as SiO₂), and sodium ions. This is essentially an irreversible process (note the single arrow) in that albite is not readily synthesized from the reaction products under ordinary natural weathering conditions, and the reaction will continue to proceed to the right as long as reactants are available.

The hydrogen ion flux that is provided by carbonic acid dissociation also can attack calcite (CaCO₃):



This reaction is relatively fast and readily reversible so that in drainage basins in carbonate-dominated terranes the stream water commonly will have near-equilibrium concentrations of hydrogen, bicarbonate, and calcium ions. At equilibrium, the rates of forward and reverse processes represented in equation 5 are equal.

In effect, the forward progress of reactions such as is shown in equations 4 and 5 will be controlled by the availability of hydrogen ions, but the final result, as indicated by the chemical composition of stream water from any given drainage basin, will be influenced by a complicated set of interrelated physical factors that influence the volume and rate of water movement, ecologic and climatic factors that control the type and density of plant and bacterial growth and soil development, and the human development of water and land resources. Thus, the geochemistry of stream water in any given drainage basin is unique to that basin. The historical record of water chemistry in a specific basin cannot be interpreted without giving proper attention to the way various hydrologic and other environmental factors in the basin have influenced the chemical composition of the water. An extensive body of research on the topic of global stream water geochemistry is summarized in Ref. 6, who agree with the need expressed in this paragraph for a broad consideration of cause and effect when evaluating stream water chemistry.

Effects of Human Activities. A considerable part of the currently existing motivation for organized long-term water quality studies has been public concern that human activities in many drainage basins have induced destructive changes in stream water quality. From examples cited in this article, such effects can indeed be documented. Also, in some basins water quality management has succeeded in correcting some of the human-caused deterioration and is substantially restoring the quality of the water. The concept of sustainability is relevant, and the development goal for drainage systems is to maintain suitable water quality while permitting levels of water use that will sustain the basin's existing and reasonable future economic development.

Human activities that alter stream flow characteristics and thus cause water quality changes include the building of structures that impound or regulate rates of stream flow, diversion of water from one drainage basin to another, irrigation of land adjacent to streams, and lowering of tributary groundwater tables by pumping from wells. Waste disposal, directly or indirectly, into streams also influences water quality by adding chemicals and suspended matter. Disposal of untreated organic waste into streams was common in urban and rural areas of the nation until the early twentieth century. Besides pathogenic bacteria in the waste, large amounts of organic chemicals and suspended material depleted the dissolved oxygen of receiving waters and killed much of the aquatic biota in some streams. Thus, the concentration of dissolved oxygen in stream water also is considered a contamination index. Normally, the dissolved-oxygen content is near the saturation level that can be calculated for water that is in contact with air at ambient temperature, but oxidizable material in solution, especially organic waste, can substantially deplete the dissolved-oxygen content. Additionally, phosphate (PO₄³⁻)

concentrations are indicative of contamination from waste sources. Phosphate is a constituent of domestic and industrial waste, in part, because of the widespread use of phosphate compounds as detergent additives.

Land use changes and related developments also can affect stream water quality. Examples include urbanization, clearing of forests, various agricultural practices, such as use of fertilizers and pesticides and return flow of drainage water from irrigated fields, and industrialization. Urbanization and industrialization lead to various side effects. Mining for coal and metals generally contaminates water during and after the mining activity. The smelting of ores to recover metals and the burning of coal to generate power release pollutants into the air, and eventually some of these pollutants find their way into water supplies. Metal ores, coal, and other organic fuels commonly contain, or are associated with, reduced sulfur. Oxidation of the sulfur by burning the fuel and smelting the ores and oxidation in the mines or in waste dumps when sulfides are exposed to air constitute major sources of sulfur in stream water. High metal concentrations commonly are found in streams draining metal-mining areas, and sodium chloride and calcium chloride are dispersed widely by salt and sand mixtures used to melt ice from highways.

Circulation Rates of Elements. The concept of cycling of individual elements, in part coupled to the hydrologic cycle, has been developed and quantified over the past half century. Besides the total quantities of the elements present in various reservoirs—bedrock (the lithosphere), soils, all forms of living matter (the biosphere), the oceans, the atmosphere, and fresh water—the rates of exchange and mechanisms of movement from one reservoir to another are considered in the cycle. This concept is highly relevant in developing a frame of reference for evaluating possible environmental effects of energy technology.

A simplified diagram representing the various reservoirs and transport mechanisms and pathways involved in the cycles of nutrient elements at and above the surface of the Earth is given in Figure 1. The processes are those considered to be the most important in the context of this article, but others of lesser significance can be postulated. For some of the elements, notably carbon, sulfur, chlorine, and nitrogen, considerable research has been done to evaluate (quantitatively) the amount of the various elements in the reservoirs and the rates of transfer.

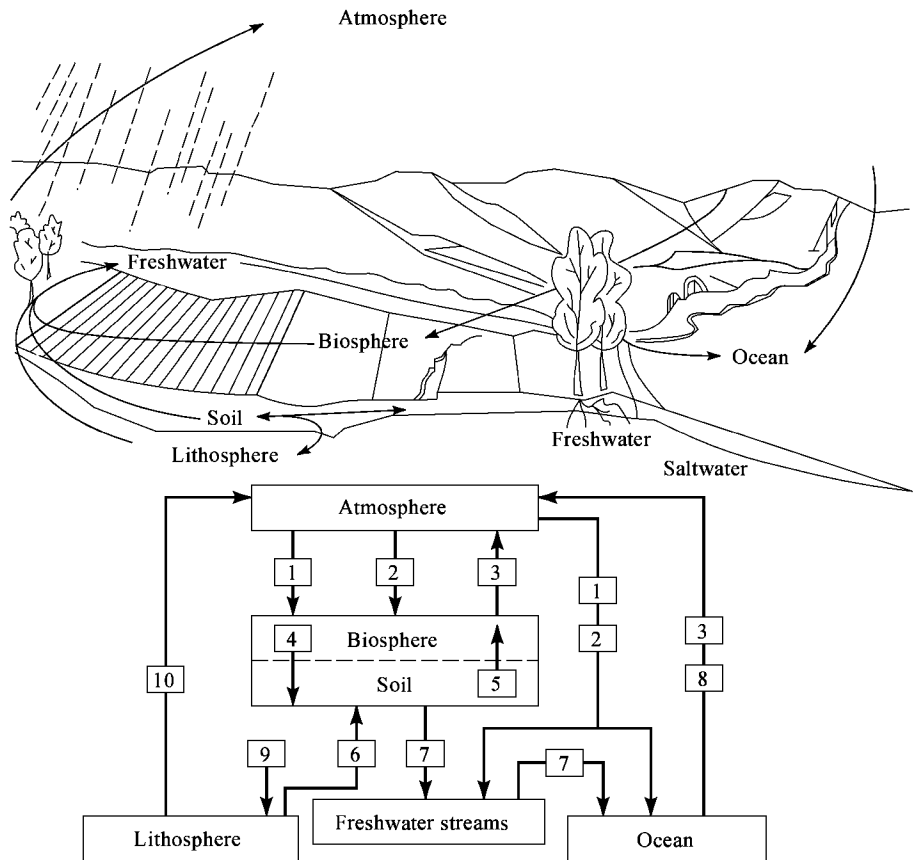


Fig. 1. Generalized cycle of the various reservoirs and transport mechanisms and pathways involved in the circulation of nutrient elements. The numbered arrows represent processes by which elements transfer among the reservoirs. Processes shown are those considered to have the most important influence on stream water quality. (Modified from Ref. 7.)

Carbon. Most of the Earth's supply of carbon is stored in carbonate rocks in the lithosphere. Normally the circulation rate for lithospheric carbon is slow compared with that of carbon between the atmosphere and biosphere. The carbon cycle has received much attention in recent years as a result of research into the possible relation between increased atmospheric carbon dioxide concentration, most of which is produced by combustion of fossil fuel, and the "greenhouse effect," or global warming. Extensive research has been done on the rate at which carbon dioxide might be converted to cellulose and other photosynthetically produced organic compounds by various forms of natural and cultivated plants. Estimates also have been made of the rate at which carbon dioxide is released to soil under optimum conditions by various kinds of plant cover, such as temperature-zone deciduous forests, cultivated farm crops, prairie grassland, and desert vegetation.

The efficiency of the weathering of rocks in using carbonic acid produced in the carbon cycle is affected by various hydrologic, environmental, and cultural controls. The fact that the principal anion in fresh surface water worldwide almost always is bicarbonate attests to the overriding importance of this process. Exceptions are systems in which evaporite minerals are available for dissolution by groundwater or where human activities are major sources of sulfate or chloride inflow.

Quantitative estimates of the magnitude of the carbon cycle, based mainly on a compilation in Ref. 8, suggest that the total amount of carbon dioxide converted to organic matter during a year might be as great as 7–9 tons/acre in midlatitude deciduous forests and about a tenth that amount in native grassland. Cultivated farm crops presumably convert somewhat more carbon dioxide per unit area than might be expected in native grassland. Respiration of carbon dioxide by plant roots is at a rate equal to about 25% of the total carbon dioxide used by the plants, and the potential supply of carbon dioxide from vegetation for weathering rock is about 1.65 tons/acre/yr on the average, over the whole land area of the Earth (8). This is equivalent to about 1060 tons/mi²/yr of carbon dioxide. If all this carbon dioxide were dissolved in water and the resulting carbonic acid reacted with silicate minerals to maintain near-neutral pH, the annual bicarbonate yield could be as great as 1500 tons/mi². A drainage basin in a limestone terrane theoretically could produce bicarbonate at an even higher rate, as the reaction between dissolved carbon dioxide and solid carbonate minerals produces two bicarbonate ions per reacting dissolved carbon dioxide molecule.

As might be expected, the average rates of bicarbonate yield in drainage basins of the world are far below these numbers, as only a relatively small proportion of the carbon dioxide released in soil is likely to participate in rock weathering. A general indication of the role of dissolved carbon dioxide in weathering processes in streams can be obtained by calculating the average annual bicarbonate yield from average concentration of dissolved bicarbonate,

average water discharge, and drainage basin area. Examples of this type of calculation can be obtained from published literature. Ref. 9 studied 56 drainage basins in the conterminous United States and Hawaii to determine the relative importance of stream temperature, human population density, atmospheric precipitation, and predominant type of bedrock (limestone, sandstone, or crystalline) as controls of constituent yield. Of the 19 limestone basins, the Kalamazoo River at Saugatuck, Michigan, had the maximum annual bicarbonate yield, which was about 324 tons mi². However, in the other basins, annual bicarbonate yields were substantially lower—15 had less than 225 tons mi², and the average yield for the 19 limestone basins was 114 tons mi².

In general, bicarbonate yields reported in Ref. 9 were much smaller for basins in crystalline (noncarbonate) rock terranes than for those in limestone or sandstone terranes. Other investigators have reported relatively large bicarbonate yields for certain igneous terranes. For example, Ref. 10 reports large bicarbonate yields from igneous drainage basins in the Pacific Northwest. Among studies referred to is a study for a basin having an active glacier in the northern Cascade Mountains of Washington (11). In this basin, the annual bicarbonate yield was near or slightly greater than 100 tons mi².

The high bicarbonate yield in some of the Pacific Northwest streams could be the result of glacial action and mechanical erosion of rock material in basins having steep slopes that produce large amounts of relatively freshly fractured and fine-grained sediment that would participate readily in dissolution and alteration reactions. The rate at which such reactions would proceed is a function of net surface area of sediment available per unit volume of runoff and the length of time such material was exposed to water in the runoff process. Therefore, the efficiency of solvent erosion under conditions in the Pacific Northwest region could be considerably increased by such factors.

The chemical analyses tabulated in this article identify "alkalinity" as a property of the water rather than a simple constituent. Alkalinity has been more broadly defined as "capacity for acid neutralization" (12,13). Common practice in water analysis is to report alkalinity in terms of bicarbonate and carbonate concentrations, although other ionic species also may contribute by reacting with the titrating acid.

Sulfur. The cycle of sulfur in weathering environments is affected by a more diverse set of reactions than for carbon. As is the case with carbon, most of the Earth's supply of sulfur is stored in the lithosphere (5). Although some sulfur actually is taken up by vegetation, the growth and decay of plants does not tie up large proportions of the total sulfur supply. The element's geochemical behavior in the environment is summarized in Ref. 14. In igneous and metamorphic rocks, sulfur generally is present in the chemically reduced sulfide form (S²⁻) and commonly is associated with metals. Polysulfide minerals in which the nominal valence of sulfur is between -2 and 0 also are common. An example is pyrite (FeS₂), which commonly occurs in association with coal seams and other sediments laid down under conditions where free oxygen was not available.

In weathering environments where oxygen is continuously available, negatively charged reduced forms of sulfur are converted to positively charged oxidized forms such as sulfate (SO₄²⁻) in which the sulfur is in the fully oxidized (S⁶⁺) state. Oxidation and reduction reactions of sulfur commonly are bacterially mediated. The calcium sulfate minerals gypsum (CaSO₄·2H₂O) and anhydrite (CaSO₄) are common constituents of evaporite rocks, and in semiarid regions where such rocks are near the land surface, stream water can contain substantial concentrations of sulfate. However, streams in humid regions generally carry relatively low concentrations of sulfate unless human activities have intervened. Weathering reactions that involve oxygen are important in the development of soils, especially in humid and subhumid climates. Commonly, oxygen is in relatively short supply at shallow depths in the soil zone, having been depleted by oxidation of organic matter. In wetland soil and submerged sediments, oxygen can be in very short supply and sulfur is in reduced form. However, as noted in Refs. 15 and 16, the sulfur in aerated soil is principally a component of organic compounds. In any event, any reduced species of sulfur may be converted to sulfate by oxidation when wetlands are drained or soils are converted from their natural state to agricultural cropland. All these processes can increase the availability of sulfate for transport by streams and cause increased sulfate loads in areas affected by development.

Geochemical studies of sulfate in streams have been approached in various ways. For example, Ref. 17 is a review of literature on sources of sulfate in the dissolved loads of streams, especially in areas where development effects are likely to be strong. It was concluded that for North America as much as 60% of the average yield of sulfate carried to the ocean is related to human activities. This study assigned a higher proportion of the total yield of sulfate in European streams to human sources, but for other continents it was thought the effects of human activities to be relatively minor.

Some of the compounds produced in the sulfur cycle are gases. For example, combustion of fossil fuels, especially coal, produces sulfur dioxide gas (SO₂), which is further oxidized in the atmosphere to sulfur trioxide (SO₃) that combines with water to form sulfuric acid (H₂SO₄). Reduction of sulfate in anaerobic soils and sediment produces hydrogen sulfide (H₂S) gas that also is reoxidized to sulfate in the presence of air. As a result, precipitation from the atmosphere is a major source of sulfate in streams in parts of North America and Europe. The estimate for human sources includes both sulfur dioxide and hydrogen sulfide reaction pathways (17). A compilation by the U.S. Environmental Protection Agency (18), summarizing research on acidic deposition, quoted estimates and measurements of "excess sulfate" yields in precipitation and in lakes and streams in the northeastern United States and southeastern Canada. Among measured data cited was a wet-plus-dry annual sulfate deposition rate for a site in the Adirondacks that ranged from about 8.2 to 16 tons mi². Although a wide range of values was observed, these rates probably can be considered reasonable for much of the eastern half of the United States. The term "excess sulfate" is defined as the amount "over and above that supplied by sea salt cycling." Ref. 9 estimates that, on average, 60% of the sulfate yield observed for the stream basins he studied was assignable to atmospheric sources.

Another source of sulfur in the global hydrologic cycle that is not related to human activities is gaseous emission of hydrogen sulfide and sulfur dioxide from volcanoes and other geothermal sources. Although effects of these emissions can be locally intense, they generally are thought to be much less significant on a global scale than the human sources.

Sulfate concentration in streams and changes over time are discussed later in this article as one of the principal indices of human influences on stream water composition. Also, it will be shown that differences and similarities in sulfate yields help in attaining a reasonable perspective on the importance of various hydrologic and geochemical characteristics of individual drainage systems.

Chlorine. Nearly all chlorine compounds are readily soluble in water. As a result, the major reservoir for this element in Figure 1 is the ocean (5). Chloride, as noted earlier, is naturally present at low levels in rain and snow, especially over and near the oceans. Widespread increases in chloride concentration in runoff in much of the United States can be attributed to the extensive use of sodium chloride and calcium chloride for deicing of streets and highways. Ref. 19 points out the importance of the increased use of deicing salt as a cause of increased chloride concentrations in streams of the northeastern United States and the role of this factor in the chloride trends in Lake Ontario. Increases in chloride concentration also can occur as a result of disposal of sewage, oil field brines, and various kinds of industrial waste. Thus, chloride concentration trends also can be considered as an index of the alternation of streamwater chemistry by human development in the industrialized sections of the world. Although chlorine is an essential element for animal nutrition, it is of less importance for other life forms.

Nitrogen. About three-fourths of the Earth's nitrogen is present in the atmosphere as nitrogen gas. Because of its importance as an essential element in plant and animal nutrition, nitrogen, in its various oxidation states and its yield and concentration, is of considerable interest in studies of human influences on stream water composition. Certain small- and medium-sized streams in the intensively developed agricultural areas of the United States have been strongly affected by nitrogen-bearing runoff from fertilized soil. However, because of its use by aquatic vegetation, the amount and form of nitrogen tend to be seasonally variable, and especially for larger streams, it is more difficult to use than sulfate or chloride as an index of human effects on water composition. Although nitrate concentrations are reported in some of the streams, no attempt at interpretation of the nitrogen chemistry is made here; however, Ref. 20, using mainly U.S. Geological Survey (USGS) data, evaluates nitrogen trends since 1905 in the Mississippi River at St. Francisville, Louisiana.

Stream Water Quality Trends

Data Available. The first comprehensive nationwide study of the chemical composition of surface water in the United States was begun by the USGS in 1905. This program received cooperative support from state agencies in Illinois, Minnesota, Kansas, California, Oregon, and Washington and entailed collection of daily samples of water at or near gaging station sites on principal streams for a period of about one year. The daily samples were combined into 10-day composites, which were filtered to remove suspended material and analyzed for principal dissolved constituents using standard "wet-chemical" procedures. As a result of decreased funding and related factors, this work was substantially curtailed after 1907; however, the studies did produce compilations of analyses for about 100 stream-sampling sites east of the 100th Meridian (21) and for about another 55 sites in the western part of the country (22). Somewhat more detailed reports for each of the six cooperating states also were issued in the USGS Water-Supply Paper series, and

Clarke (23) used many of the analyses from the program in his summary of the composition of river and lake waters of the United States. From about 1914 to 1940 studies of river water quality were carried on at only a few sites in the United States. A revival of interest began in the 1940s, however, and by the 1960s, a much larger number of sampling sites were in operation than had existed at the height of the 1905–1907 program. The available historical data base for most U.S. sampling stations has no water quality measurements for the 1910–1940 period.

Stream water quality commonly varies greatly in response to water discharge; thus, a single year of record is not adequate for reliable extrapolation, and in any exacting comparison of historical data, this factor needs to be taken into account. From the beginning, it has been a general policy in the USGS surface water quality program to locate sampling sites at or near gaging stations where records of stream flow are obtained. Until about 1970, many of the USGS water quality records were based on daily sampling, generally with determinations of specific electrical conductance (an indicator of total cation and anion concentration) on each sample, but with extensive analyses performed only on composited daily samples. Composites usually contained 10 daily samples. However, where stream discharge and other factors caused substantial day-to-day changes in specific conductance, the composite period was shortened to prevent mixing of chemically dissimilar samples and to give a clearer indication of the stream chemistry variability. Annual averages of these analyses, weighted by time or discharge, were used to summarize the records. After 1970, complete analyses were done on single samples collected at various time intervals ranging from semimonthly to quarterly, and analyses of composite samples were no longer made. Analytical procedures changed from time to time as improved instrumentation and techniques became available.

In order to identify and consider the importance of the various sources and controlling factors that operate in specific drainage basins in the United States to produce water having hydrochemical properties that are characteristic of each basin, four drainage basins having long-term hydrologic and water chemistry records were chosen for study in the original version of this article (24)—the Great Lakes—Upper St. Lawrence River at and near Ogdensburg, New York, and Cornwall, Ontario, Canada; the Columbia River at and upstream from the Dalles, Oregon; the Allegheny River upstream from Pittsburgh, Pennsylvania; and the lower Mississippi River at and upstream from New Orleans, Louisiana. However, these discussions are beyond the scope of this article. The Great Lakes study is given here as an example. For more details, refer to Ref. 24.

The historical records mentioned above are compared with more recent data for each of the drainage basins in order to detect and explain major differences between water composition observed early in the twentieth century and that observed more recently. Sulfate, as mentioned earlier, is one of the principal indicators of the effects of human activities on stream water composition, and it is the principal indicator used in the following discussions. The records selected for study display general trends in concentrations and yields that appear to be well-enough defined to outweigh the influences of different sampling frequencies and changing compositing practices. Trends that can be detected in the data collected in the past 20 years generally are too subtle to be evaluated closely by the methods used in this article and are more appropriately studied by the more sophisticated procedures which are described in Refs. 3, 25, and 26.

Selected analytical and related data for the example drainage basin are given in Tables 1 and 2. Apparent trends with time are evaluated by various means described in the text. The annual minimum and average constituent concentrations and annual yields of sulfate in tons per unit area of drainage basin are the focus of the discussion.

Table 1. Chemical Composition of Water and Related Data for St. Lawrence River, Representing Outflow from Lake Ontario, Selected Years (1906–1990)

Constituent property, and related data	Ogdensburg, N.Y.	Cornwall, Ontario, Canada		Ogdensburg, N.Y.		Cornwall, Ontario, Canada		
	1906–1907 ^a (1)	1935 ^b (2)	1940 ^b (3)	1956 ^c (4)	1969 ^d (5)	1977 ^b (6)	1980 ^b (7)	1990 ^e (8)
silica (SiO ₂)	6.6	6.9	4.1	3.5	0.6	0.2	0.5	
calcium (Ca)	31	33	36	36	39	38	36	
magnesium (Mg)	7.2	7.9	8.5	8.2	7.3	8.0	7.8	
sodium (Na)				9.7	12	13	13	
potassium (K)	{6.3 ^f	7.3 ^f	7.5 ^f	1.5	1.3	1.6	1.4	
alkalinity as bicar-bonate (HCO ₃) (property)	122	110	97	112	110	110	99	
sulfate (SO ₄)	12	19	21	25	28	27	26	27
chloride (Cl)	7.7	15	16	21	26	27	26	
fluoride (F)				0.1	0.1	0.1	0.2	
nitrate (NO ₃)	0.3	0.6	1.0	1.0	0.3	0.53	0.49	
dissolved solids	134	156	158	179	169	169 ^g	160 ^h	
specific conductance, μS at 25°C (property)				299	314	370	270	300
pH units (property)		8.0	8.0	6.8–7.9	7.3–8.0	7.8	7.1	
drainage area, mi ²	295,200	298,800	298,800	295,200	295,200	298,800	298,800	298,800
average discharge, ft ³ s	242,300	192,000	226,000	257,000	270,000	250,000	296,000	264,800

Note: Concentration values are in milligrams per liter unless otherwise noted. Data from the Ogdensburg, N.Y., and Cornwall, Ontario, Canada, stations are considered equivalent. *Sources:* column 1, ref. 21; columns 2, 3, ref. 36; column 4, ref. 34; column 5, ref. 35; columns 6, 7, ref. 2; column 8: U.S. Geological Survey, New York District office written communication, 1991.

^a Average of 11 monthly samples, Sept. 1906–Jan. 1907, March–Aug. 1907.
^b Single sample collected on Aug. 30, 1935 (col. 2), Aug. 22, 1940 (col. 3), June 27, 1977 (col. 6), and Sept. 29, 1980 (col. 7).
^c Average of 10-day composite samples collected during water year 1969.
^d Average of 10 individual monthly samples collected during water year 1969.
^e Average of 5 bimonthly samples collected between October 1989 and July 1990.
^f Sodium and potassium not determined separately; value reported is total Na + K expressed as an equivalent amount of sodium.
^g Maximum total dissolved solids for period 1977–1980.
^h Minimum total dissolved solids for period 1977–1980.

Table 2. Estimated Annual Yield of Constituents (t/mi²), St. Lawrence River Outflow from Lake Ontario, Selected Years (1906–1990)

Constituent or property	Ogdensburg, N.Y., 1906–1907 (1)	Cornwall, Ontario, (2)	Ogdensburg, N.Y.		Cornwall, Ontario, Canada		
			1956 (5)	1969 (6)	1977 (7)	1980 (8)	1990 (9)

Canada, 1937							
(3)							
calcium (Ca)	25.0	24.2	30.9	35.1	31.3	35.1	na
sodium (Na)	5.1 ^a	6.7 ^a	8.3	10.8	10.7	12.7	na
bicarbonate (HCO ₃)	98.6	75.9	96.0	99.0	90.6	96.5	na
sulfate (SO ₄)	9.7	14.8	21.4	25.2	22.2	25.4	23.6
chloride (Cl)	6.2	10.8	18.0	23.4	22.2	25.4	na

Note: Calculated from data in Table 1 using the equation
tons per square miles annual averageconstituent concentration in milligrams per liter × annual average stream discharge in cubic feet
per second × conversion factor 0.9844 divided by drainage area in square miles . Column numbers are those used in Table 1. na, data not available for
calculation.

^a Equivalent amount of sodium from combined sodium and potassium.

Other Constituents of Stream Water. The records reported in Refs. 21 and 22 were obtained for the primary purpose of evaluating the suitability of surface water resources of the United States for utilization by industry and for irrigation of agricultural lands in the western part of the country. These stream waters also provide public water supplies for many municipalities. Evaluations of water quality for the latter purpose emphasize constituents that were not given detailed consideration in Refs. 21 and 22 summaries, although there are references in Ref. 21 to work done in various state health laboratories and municipal treatment plants.

Surface Water Ecology. Ecologic studies of contaminated and uncontaminated streams have been the subject of extensive research. Those streams that have been subjected to contaminant additions through disposal of industrial and other forms of waste generally show substantial decreases in the diversity and abundance of instream biota. These effects are often first observed in decreases of native game fish such as brook or rainbow trout, but more extensive research has shown many more subtle changes. A compilation (27) includes summaries of ecologic studies in 12 river basins in North America.

The drainage basin of the Clark Fork, an eastern tributary of the Columbia River, has been affected for long distances by the copper mining and ore treatment processes that were conducted at Butte and Anaconda, Montana, for about 125 years, ending in 1982. Conditions in this stream have been described (28). A more recent paper (29) pointed out that many streams in mineralized areas had relatively high metal concentration prior to the beginning of mining activity.

Means for Summarizing Dissolved-Material Contents. Water analyses traditionally have included a summarizing value termed "total dissolved solids." A determined value specified as "residue on evaporation" (ROE) is obtained by evaporating a measured volume of the water sample to dryness in a dish whose weight is known exactly, drying the residue at a specified temperature (in USGS practice at 180°C for 1 h), and weighing the dish and its contents after cooling in a moisture-free atmosphere. The more modern analyses also include electrical conductance, which is related to the total concentration of dissolved cations and anions, and electronic determination of pH, which is a measure of the acidity or hydrogen ion (H⁺) activity. Although the ROE procedure appears to be simple and direct, it has complicating factors. The bicarbonate in solution in the original sample is converted to carbonate by drying, with loss of an equivalent amount of carbon dioxide gas. However, some types of water, such as those having high concentrations of dissolved calcium and sulfate, can deposit residues that contain water incorporated in crystalline minerals such as gypsum (CaSO₄·2H₂O). To avoid this problem, a calculated total dissolved solids (SUM) commonly is reported for water containing more than about 1000 mg L total constituent concentration. The SUM calculation is made by adding reported concentrations for all the major constituents, with a correction factor converting dissolved bicarbonate to an equivalent carbonate content of the dried residue.

Example: Great Lakes–Upper St. Lawrence River

The St. Lawrence River, at the outlet of Lake Ontario, represents the drainage from the Great Lakes basin and has a relatively constant chemical composition, owing to the large storage capacity of the lakes and their efficiency in mixing of inflow. The water surface of the Great Lakes makes up a substantial fraction of the basin area. The basin also has a relatively constant water discharge. These characteristics make it a good site for study of water quality trends. Data from two water-quality-sampling and stream-flow-gaging stations—Cornwall, Ontario, Canada, and Ogdensburg, New York (40 miles upstream from Cornwall)—are used to show constituent change over time. Although the two stations are 40 miles apart and there is a small amount of tributary inflow between them, the chemical composition of the water at the sites does not differ significantly and thus can be used for comparison of data over time. At Cornwall, the drainage area of the St. Lawrence River is 298,800 mi² and the long-term average discharge is 245,000 ft³ s, as reported by the U.S. Geological Survey (30).

Data from these stations are summarized briefly in Table 1, where each of the columns represents data for a particular year progressing from 1906–1907 to 1990. The data show that principal dissolved-ion concentrations in the river were nearly constant during the period 1969–1980 and that most of the change occurred between 1907 and 1956. Data for some major ions for samples from Lake Ontario that were presented (31) indicate that relatively continuous increases occurred until the 1960s. An illustration in Ref. 2 based on a graph from Ref. 32 suggests that sulfate concentration in Lake Ontario steadily increased from 12–14 mg L in 1906 to 28–30 mg L in the 1970s. However, during the 1980s, the sulfate concentration in the lake as represented by the St. Lawrence River at Cornwall did not show any further upward trend, and it seems likely that the sulfate concentration had stabilized. Samples taken at 1- or 2-month intervals during water years 1981–1985 had sulfate concentrations of 28 mg L or more on 11 occasions. During water years 1986–1990 no sulfate concentrations greater than 27 mg L were observed at Cornwall.

Behavior of chloride concentration during the period of record was somewhat similar to that of sulfate—about a threefold increase compared to the doubling of sulfate. Data in Table 1 indicate that a long-term decrease in bicarbonate of 10–20 mg L is a reasonable estimate.

Refs. 32 and 33 attribute the historical changes in water composition of the Great Lakes to human activities, which have increased as the population of the drainage basins of the lower four lakes has increased since the beginning of the twentieth century. Industrial waste and incompletely treated sewage are thought to be major sources of sulfate and chloride. Other nonpoint sources are agricultural fertilizers and fossil fuel combustion and urban runoff that add sulfate and sodium and calcium chlorides used for deicing highways. The decrease in alkalinity of the water leaving Lake Ontario can possibly be explained as a direct consequence of the low pH of rain that falls on all the lakes. However, because the loss of bicarbonate alkalinity is only about two-thirds as great as the gain in sulfate, the observed changes cannot be simply assigned to sulfuric acid in this rainfall. It seems more likely that some of the acidity in the rain that falls over the tributary drainage is neutralized by reaction with sedimentary rock minerals, as calcium and sodium show increases.

During the 1960s there was much concern about the increasing rate of eutrophication of Lake Erie. The limiting nutrient for aqueous microbiota in the Great Lakes was perceived to be phosphorus (33), and household synthetic detergents containing sodium phosphate were thought to be an important source. Consequently, a concerted effort was made to decrease the sale of detergents containing phosphate and to improve sewage treatment processes. By the 1980s, some declines in phosphate concentration had occurred in Lake Erie and Lake Ontario (33).

An accompanying effect of eutrophication that is more readily observable in Table 1 is a decrease in silica concentration in Lake Ontario. Some decline in dissolved silica apparently has occurred in all of the lakes except Lake Superior. This decline is brought about by the growth of diatoms, a species of aquatic microorganisms in the upper layers of lake water that is widespread in all types of water impoundments where the water is clear and exposed to the sun. The silica is used by these microorganisms to form their skeletons and is later precipitated and becomes part of the bed sediment.

Silica determinations on the 1906–1907 St. Lawrence samples (Table 1, col. 1) might have been affected by inadequate filtration, and possibly some silica was dissolved from the glass sample bottles during storage before analysis. Thus, the average silica concentration of 6.6 mg L might be too high. However, a rather well-defined downward trend is observable in more recent silica concentration records for the St. Lawrence at Ogdensburg and Cornwall that has greater experimental certainty. Determinations of dissolved silica on individual samples collected during water year 1956 ranged from 1.1 to 6.8

mg L (34). A record of once-monthly samples from water year 1969 (35) shows silica concentration that ranged from a low of 0.0 (below the minimum reporting limit) to a maximum of 1.9 mg L. Samples taken during water years 1977–1980 show concentrations of 0.2 and 0.5 mg L (Table 1, cols. 6, 7). The Lake Superior drainage basin is underlain predominantly by metamorphic and igneous rocks, whereas most of the Great Lakes drainage area eastward to the outlet of Lake Ontario is underlain by carbonate-rich sediment. It would be expected, therefore, that the water of Lake Superior would have different chemical characteristics than those of the other Great Lakes because it has a different set of geochemical controls. Also, the major ion composition of water in Lake Superior has not changed significantly during the period of record.

Additional insight into changes in dissolved material flowing from the lakes over time can be obtained by calculating average annual yields of major constituents observed in the Great Lakes–St. Lawrence basin (Table 2). Ref. 9 contains calculations of the annual yields of principal constituents for 56 drainage basins and evaluated the degree of correlation with four environmental factors—bedrock type, annual precipitation, population density, and average stream temperature. The study shows that in basins dominated by limestone bedrock, the average annual yield for calcium was 36.0 tons mi² and for bicarbonate was 114.2 tons mi². Calcium yields listed in Table 2 for the 1956–1980 period are close to the averages reported by Peters (9) and are significantly higher than yields for 1906–1907. However, the bicarbonate yields indicated in Table 2 are lower than the average reported by Peters and show no well-defined trend. Eight of the 56 streams studied by Peters are tributaries of the Great Lakes–St. Lawrence system, and the average of the annual sulfate yield reported for these eight streams was 28.5 tons mi². This finding agrees approximately with the sulfate yield of 25.4 tons mi² given in Table 2 (col. 8) for Cornwall during water year 1980, considering the rather wide range of annual sulfate yields in the eight basins that were selected.

Although the substantial effects of human activities on the composition of the Great Lakes water is well documented by the data cited here, the system as a whole in recent years has shown a substantial ability to maintain a relatively high degree of chemical stability. The effect of acidic precipitation, for example, has not significantly depleted the capacity of the system to neutralize acid as has occurred in smaller lakes and streams in noncarbonate terranes in other parts of the northeastern United States. Although further increases in sulfate concentration in the lower Great Lakes have been predicted by various investigators (33), the concentrations of sulfate and chloride in the St. Lawrence River just downstream from the outlet of Lake Ontario appear to have remained the same or perhaps declined slightly since about 1980 (26).

The annual yields of sulfate during 1906–1990 for the St. Lawrence–Great Lakes basin (Fig. 2) are influenced considerably by differences in water discharge from year to year. The maximum values of about 25 tons mi² in 1969 and 1980 and 23 tons mi² in 1990 represent discharges substantially above the long-term average. The minimum yield for 1960–1970, 18.8 tons mi², occurred in 1966, a year that had below-normal discharge. However, the yield more than doubles between 1906 and 1956 and has an upward trend through about 1970.

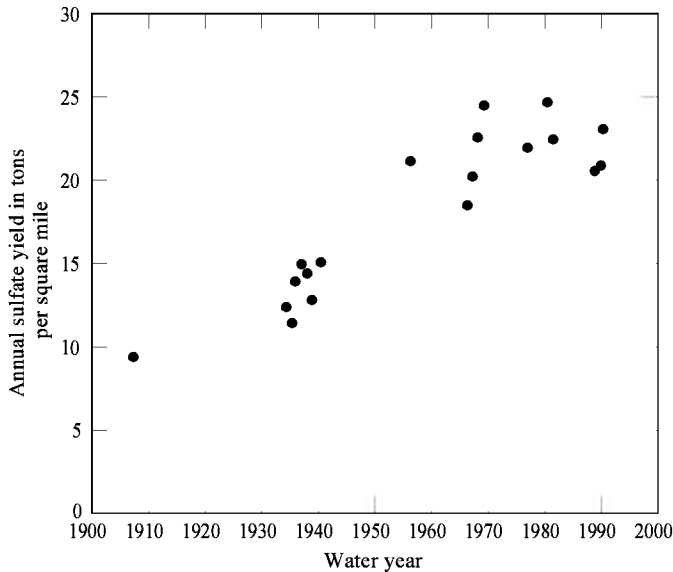


Fig. 2. Annual sulfate yields of the St. Lawrence River at Cornwall, Ontario, Canada, 1906–1990. (Calculated from data in U.S. Geological Survey water-resources records and Ref. 36.)

Summary

Under pristine conditions, that is, in the absence of human civilization and development, the chemical composition of inland stream and lake waters is, ideally, controlled by the alteration of rock minerals through chemical weathering processes, which liberate soluble products. These processes in turn are controlled or influenced by climatic factors such as rainfall, air temperature, and evaporation and by associated biological or biochemical processes, such as photosynthesis and transpiration by plants, decay of vegetative debris, and the effects of aquatic life processes. Circulation of essential nutrient elements, including carbon, sulfur, chlorine, and nitrogen, generally is bound to elemental oxygen from the atmosphere and provides most anionic species occurring in natural water, such as bicarbonate, sulfate, chloride, and nitrate. Other constituents of natural surface waters, including calcium, magnesium, sodium, and potassium, can be correlated in general with the chemical composition of rocks and soils in a given drainage basin and are found as principal cationic species and are in electrochemical balance with anions in these waters.

The influence of human activities in a stream drainage basin can be relatively simple and direct, as in the disposal of soluble organic and inorganic waste, or more subtle and complex, as in the conversion of prairie or forest land to agricultural use. Such effects can be expected to increase as population density and agricultural, industrial, and mining activities increase.

In the Great Lakes–Upper St. Lawrence River basin, for example, the yield of sulfate in tons per square mile per year in the St. Lawrence River nearly doubled between 1905 and 1956 and continued to increase, but at a lesser rate, until about 1970, when the yield leveled off or perhaps even declined slightly. The continuing yield of between about 19 and 25 tons mi² indicates that the basin may have reached a steady state between the natural and human-induced loading of sulfur to the basin and its removal by the St. Lawrence River. In contrast, sulfate concentrations in the upper Columbia River basin at Northport, Washington, show less clearly defined trends in sulfate concentrations and yields during the century (24). These data indicate that human-induced effects are largely masked by the large amount of runoff available in the Columbia River basin and by the effects of storage and mixing in lakes and reservoirs.

Coal mining was extensive in the Allegheny River drainage basin in Pennsylvania in this century, and sulfate concentrations in the river near Pittsburgh increased substantially between the early 1900s and 1962 as a result of drainage from many active and abandoned coal mines. The operation of flood control and multipurpose reservoirs in the basin has caused a nearly 50% decline in sulfate concentration in the Allegheny River near Pittsburgh at low flow, as shown in analysis of samples collected in 1947 and compared to samples collected in 1975 and 1989 (24). This flow augmentation by timed reservoir releases has had a beneficial effect on the quality of the Allegheny River.

The Mississippi River drains more than 1,125,000 mi² of the conterminous United States and integrates the effect on stream water quality of a large

range of human activities across a large continental area. The calculation of sulfate yield in the basin has many uncertainties because water quantity and water quality data are incomplete, and the effects of flow control and water diversions are difficult to measure. However, from the available data it can be estimated that sulfate concentrations in the lower Mississippi River at and upstream from New Orleans, Louisiana, probably have doubled between 1905–1906 and 1989; most of this increase seems to have occurred before 1980 (24). Estimates of the increase in annual sulfate yield due to human activities since 1905 ranging from about 9 to about 14 tons mi² are consistent with the increase in sulfate yields in the Great Lakes–St. Lawrence River system. In both instances, yields seem to have leveled off around 1970 or 1980 and to have remained fairly stable since. Possibly, both drainage systems have reached a steady state, and the natural and human-induced sulfate loading to the basins, much of it from atmospheric deposition, is now stable. Data for total SO₂ emissions for each state from 1965 to 1980 (37) show that in most states adjacent to the Great Lakes the emission rates decreased significantly after 1970.

BIBLIOGRAPHY

"Water, Sources and Utilization" in *ECT* 2nd ed., Vol. 21, pp. 652–668, by R. G. Dressler, Trinity University; "Water, Sources and Utilization" in *ECT* 3rd ed., Vol. 24, pp. 257–275, by Gerald Meyer, U.S. Geological Survey.

1. American Water Works Association, American Public Health Association, and Water Environment Foundation, *Standard Methods for the Examination of Water and Wastewater*, 18th ed., American Public Health Assc., Washington, D.C., 1992.
2. J. D. Hem, A. Demayo, and R. A. Smith, in *Surface Water Hydrology*, V O-1, *The Geology of North America*. The Geological Society of America, Boulder, Colo., 1990, pp. 189–231.
3. R. A. Smith, R. B. Alexander, and M. G. Wolman, *Analysis and Interpretation of Water-Quality Trends in Major U.S. Rivers*, 1974–81, U.S. Geological Survey Water-Supply Paper 2307, U.S. Geological Survey, Reston, Va., 1987.
4. V. M. Goldschmidt, *Geochemistry*, Clarendon, Oxford, England, 1954.
5. J. D. Hem, *Study and Interpretation of the Chemical Characteristics of Natural Water*, 3rd ed., U.S. Geological Survey Water-Supply Paper 2254, U.S. Geological Survey, Reston, Va., 1985.
6. M. Meybeck and R. Helmer, *Palaeogeogr. Palaeoclimatol. Palaeoecol. (Global Planet. Change Sec.)* **75**, 283–309 (1989).
7. M. Meybeck, *Int. Assoc. Hydrol. Sci. Bull.* **141**, 173–192 (1983).
8. H. Lieth, *J. Geophys. Res.* **68**(13), 3887–3898 (1963).
9. N. E. Peters, *Evaluation of Environmental Factors Affecting Yields of Major Dissolved Ions of Streams in the United States*, U.S. Geological Survey Water-Supply Paper 2228, U.S. Geological Survey, Reston, Va., 1984.
10. D. P. Dethier, "Weathering Rates and the Chemical Flux from Catchments in the Pacific Northwest, U.S.A.," in S. M. Colman and D. P. Dethier, eds., *Rates of Chemical Weathering of Rocks and Minerals*, Academic, Orlando, Fla., 1986, pp. 503–530.
11. R. C. Reynolds, Jr. and N. M. Johnson, *Geochim. Cosmochim. Acta* **36**, 537–554 (1972).
12. J. R. Kramer, "Alkalinity and Acidity," in R. A. Minear and L. H. Keith, eds., *Water Analysis*, vol. 1: *Inorganic Species*, part 1. Academic Press, New York, 1982, pp. 85–135.
13. W. Stumm and J. J. Morgan, *Aquatic Chemistry*, 2nd ed., Wiley-Interscience, New York, 1981.
14. J. O. Nriagu, ed., *Sulfur in the Environment, Part II: Ecological Impacts*, Wiley-Interscience, New York, 1978.
15. M. Nyborg, "Sulfur Pollution in Soils," in J. O. Nriagu, ed., *Sulfur in the Environment, Part II: Ecological Impacts*, Wiley-Interscience, New York, 1978, pp. 359–390.
16. J. W. Fitzgerald, "Naturally Occurring Organosulfur Compounds in Soil," in J. O. Nriagu, ed., *Sulfur in the Environment, Part II: Ecological Impacts*, Wiley-Interscience, New York, 1978, pp. 391–443.
17. R. A. Berner, *J. Geophys. Res.* **76**, 6597–6600 (1971).
18. U.S. Environmental Protection Agency (EPA), *The Acidic Deposition Phenomenon and Its Effect*, vol. 2: *Effects Sciences*, Critical Assessment Review Papers, Report EPA 600/9-83-016AF, EPA, Washington, D.C., 1984, pp. 4–11.
19. R. C. Bubeck, W. H. Diment, B. L. Deck, A. L. Baldwin, and S. D. Lipton, *Science* **172**, 1128–1132 (1971).
20. R. E. Turner and N. N. Rabalais, *Bio-Science* **41**(3), 140–147 (1991).
21. R. B. Dole, *The Quality of Surface Waters of the United States—Part 1, Analyses of Waters East of the One-Hundredth Meridian*, U.S. Geological Survey Water-Supply Paper 236, U.S. Geological Survey, Washington, D.C., 1909.
22. H. Stabler, *Some Stream Waters of the Western United States with Chapters on Sediment Carried by the Rio Grande and the Industrial Application of Water Analyses*, U.S. Geological Survey Water-Supply Paper 274, U.S. Geological Survey, Washington, D.C., 1911.
23. F. W. Clarke, *The Composition of the River and Lake Waters of the United States*, U.S. Geological Survey Professional Paper 135, U.S. Geological Survey, Washington, D.C., 1924.
24. J. D. Hem, "Water Quality Issues," in A. Bisio and S. Boots, eds., *Encyclopedia of Energy Technology and the Environment*, Vol. 4, John Wiley & Sons, Inc., New York, 1995, pp. 2799–2823.
25. D. R. Helsel, "Statistical Analyses of Water Quality Data," in R. W. Paulson, E. B. Chase, J. S. Williams, and D. W. Moody, eds., *Compilers, National Water Summary 1990–91*, U.S. Geological Survey Water-Supply Paper 2400, U.S. Geological Survey, Reston, Va., 1993, pp. 93–100.
26. R. A. Smith, R. B. Alexander, and K. J. Lanfear, "Stream Water in the Conterminous United States—Status and Trends of Selected Indicators During the 1980's," in R. W. Paulson, E. B. Chase, J. S. Williams, and D. W. Moody, eds., *Compilers, National Water Summary 1990–91*, U.S. Geological Survey Water-Supply Paper 2400, U.S. Geological Survey, Reston, Va., 1993, pp. 111–140.
27. C. D. Becker and D. A. Neitzel, eds., *Water Quality in North American River Systems*, Battelle, Columbus, Ohio, 1992.
28. J. N. Moore and S. N. Luoma, *Environ. Sci. Technol.* **24**, 1278–1285 (1990).
29. D. O. Runnells, T. A. Shepherd, and E. E. Angino, *Environ. Sci. Technol.* **26**, 2316–2323 (1992).
30. U.S. Geological Survey, *Water Resources Data for New York*, vol. 1: *Eastern New York Excluding Long Island—Water Year 1988*, U.S. Geological Survey Water Data Report NY 551, Springfield, Va., National Technical Information Service 1989.
31. R. R. Weiler and V. K. Chawla, "Dissolved Mineral Quality of Great Lakes Waters," in *Proceedings of the 12th Conference on Great Lakes Research, Ann Arbor, Mich.*, International Association for Great Lakes Research, 1969, pp. 801–818.
32. H. F. H. Dobson, "Principal Ions and Dissolved Oxygen in Lake Ontario," in *Proceedings of the 10th Conference on Great Lakes Research, Toronto, Canada, 1967, Ann Arbor, Mich.*, International Association for Great Lakes Research, 1967, pp. 337–356.
33. R. R. Weiler, "Chemistry of the North American Great Lakes," *Verhandlungen der Internationalen Vereinigung für Theoretische und Angewandte Limnologie*, **21**, 1681–1694 (1981).
34. U.S. Geological Survey (USGS), *Quality of Surface Waters of the United States 1956*, Parts 1–4, U.S. Geological Survey Water-Supply Paper 1450, USGS, Washington, D.C., 1960.
35. U.S. Geological Survey (USGS), *Quality of Surface Waters of the United States 1969*, Parts 4–5, U.S. Geological Survey Water-Supply Paper 2144, USGS, Reston, Va., 1974.
36. H. A. Leverin, *Industrial Waters of Canada—Report of Investigations 1934 to 1940*, Bureau of Mines Report 807, Canada Department of Mines and Resources, Ottawa, Ont., 1942.
37. R. A. Smith and R. B. Alexander, *Evidence for Acid-Precipitation Induced Trends in Stream Chemistry at Hydrologic Bench-Mark Stations*, U.S. Geological Survey

Circular 910, U.S. Geological Survey, Reston, Va., 1983.

John D. Hem
U.S. Geological Survey Water Resources Division

PROPERTIES

The most striking feature of the earth, and one lacking from the neighboring planets, is the extensive hydrosphere. Water is the solvent and transport medium, participant, and catalyst in nearly all chemical reactions occurring in the environment. It is a necessary condition for life and represents a necessary resource for humans. It is an extraordinarily complex substance. Structural models of liquid water depend on concepts of the electronic structure of the water molecule and the structure of ice. Hydrogen bonding between H₂O molecules has an effect on almost every physical property of liquid water.

Natural water systems contain numerous minerals and often a gas phase (Fig. 1). They include a portion of the biosphere and organisms, and their abiotic environments are interrelated and interact with each other. The distribution of chemical species in waters is strongly influenced by an interaction of mixing cycles and biological cycles.

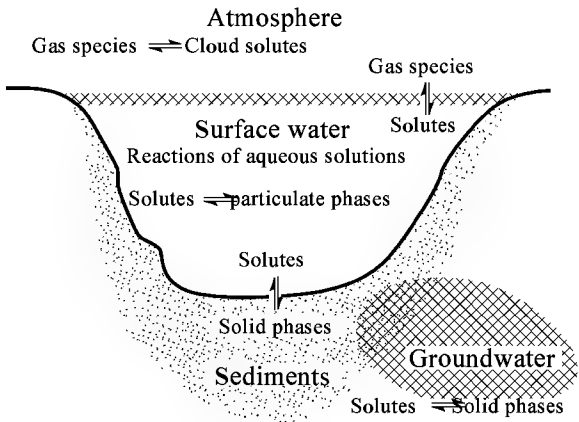


Fig. 1. Natural water environments. Water links essential cycles of the atmosphere with those of the sediments.

Human civilization interferes more and more with the cycles that connect land, water, and atmosphere, and pollution seriously affects water quality. In order to assess the stresses caused to aquatic ecosystems by chemical perturbation, the distribution of pollutants and their fate in the environment must be investigated (see AIR POLLUTION).

The Solvent Water

Water is an unusual liquid. It has a very high boiling point and high heat of vaporization; ice has a very high melting point. The maximum density of liquid water is near 4°C, not the freezing point, and water thus expands upon freezing (Fig. 2, Tables 1 and 2). It has a very high surface tension. It is an excellent solvent for salts and polar molecules. It has the greatest dielectric coefficient of any liquid. These unusual properties are a consequence of the dipolar character of the H₂O molecule. Figure 3 depicts the electron cloud of the angular water molecule, resulting from the hybridization of *s* and *p* electrons to yield two bonding orbitals between the O and the two H atoms, and two nonbonding *sp*³ orbitals on the oxygen. The molecule thus has high negative charge density near the oxygen atom and high positive charge density near the protons. It is a dipolar molecule.

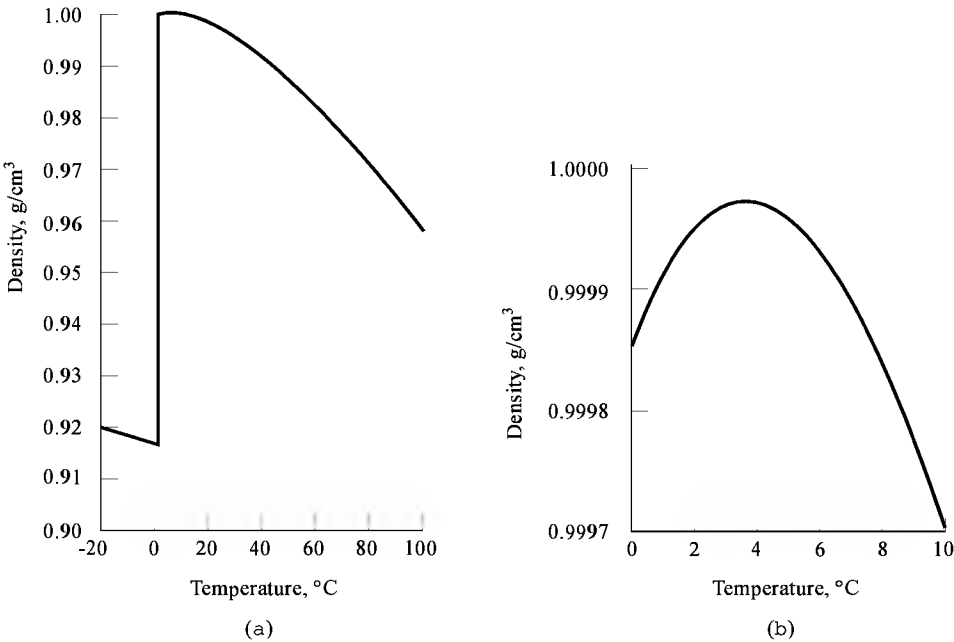


Fig. 2. (a) The density of ice and liquid water at 101.3 kPa (1 atm) as a function of temperature, and (b) the density in the domain of its maximum (1).

Table 1. Thermodynamic Constants for Phase Changes of H₂O^a

Property	Fusion ^b	Vaporization ^b	Sublimation ^c
temperature, K	273.15	373.15	273.16
isopiestic heat capacity change Δc_p , J (mol°C) ^d	37.28	41.93	
enthalpy change ΔH , kJ mol ^d	6.01	40.66	51.06
entropy change ΔS , J (mol°C) ^d	22.00	108.95	186.92
volume change ΔV , cm ³ mol	1.621	3.01×10^4	
internal energy change ΔE , kJ mol ^d	6.01	37.61	48.79

^a Ref. 1.
^b At 101.3 kPa (1 atm).
^c At ice–liquid–vapor triple point.
^d To convert J to cal, divide by 4.184.

Table 2. Physical and Chemical Properties of Liquid Water^a

Property	Comparison with other substances	Importance to environment
density	maximum density at 4°C, not at freezing point; expands upon freezing (both properties unusual)	in lakes prevents freezing up and causes seasonal stratification
melting and boiling points	abnormally high	permits water to exist as a liquid at earth's surface
heat capacity	highest of any liquid except ammonia	moderates temperature by preventing extremes
heat of vaporization	one of the highest known	important to heat transfer in atmosphere and oceans; moderates temperature extremes
surface tension	very high	regulates drop formation in clouds and rain
absorption of radiation	large in infrared and ultraviolet regions; less in visible regions	important control on biological activity (photosynthesis) in water bodies and on atmospheric temperature
solvent properties	excellent solvent for ionic salts and polar molecules because of dipolar nature	important in transfer of dissolved substances in hydrological cycle and in biological systems

^a Ref. 2.

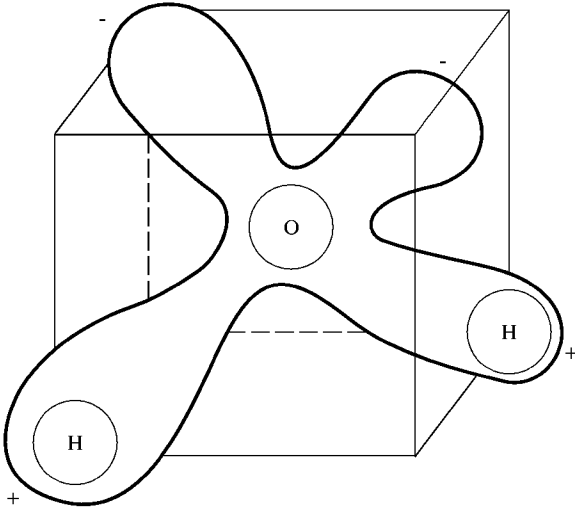


Fig. 3. Electron cloud depiction for the H₂O molecule (3).

Figure 4 shows the measured angle of 105° between the hydrogens and the direction of the dipole moment. The measured dipole moment of water is 1.844 debye (a debye unit is 3.336×10^{-30} C m). The dipole moment of water is responsible for its distinctive properties in the liquid state. The O–H bond length within the H₂O molecule is 0.96×10^{-10} m. Dipole–dipole interaction between two water molecules forms a hydrogen bond, which is electrostatic in nature. The lower part of Figure 4 (not to the same scale) shows the measured H-bond distance of 2.76×10^{-10} m or 0.276 nm.

The hydrogen-bonded structure of ice is shown in Figure 5a, and one of the several models proposed for the structure of water in the liquid state is shown in Figure 5b. In the open tetrahedral ice structure, each oxygen atom is bound to four nearby oxygen atoms by H bonds. The energy of each H bond is estimated to be about 20 kJ mol⁻¹; by comparison, covalent bond energies are typically 20 times greater. H bonds are of low energy, but they are numerous in ice and water (5).

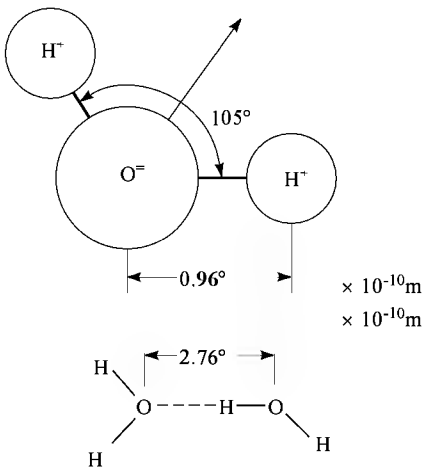


Fig. 4. Structure of the angular water molecule and the hydrogen bond (3).

Upon melting, ice loses its open structure with the "melting" of some fraction of the hydrogen bonds, and so the volume of the liquid water decreases, reaching a minimum at 4°C; above this temperature thermal expansion dominates the density.

In Figure 5b the Frank-Wen flickering cluster model envisions larger clusters of H-bonded water surrounded by noncluster waters, which nonetheless interact with neighbors by dipole–dipole forces. The lifetime of the clusters is estimated at around 100 ps, which is long with respect to the period of a molecular vibration of approximately 0.1 ps. The persistence, or the reformation, of hydrogen bonding in liquid water is a key to understanding the physical properties of water as well as its poorer solvent properties for nonpolar, hydrophobic solutes. The highly structured water linked by H bonds must be disrupted by any solute. When the solute is ionic, the attractive interactions between ion and water molecule favor dissolution. When the solute is nonpolar molecule, the structural cost to the hydrogen-bonded water makes the probability of dissolution unfavorable.

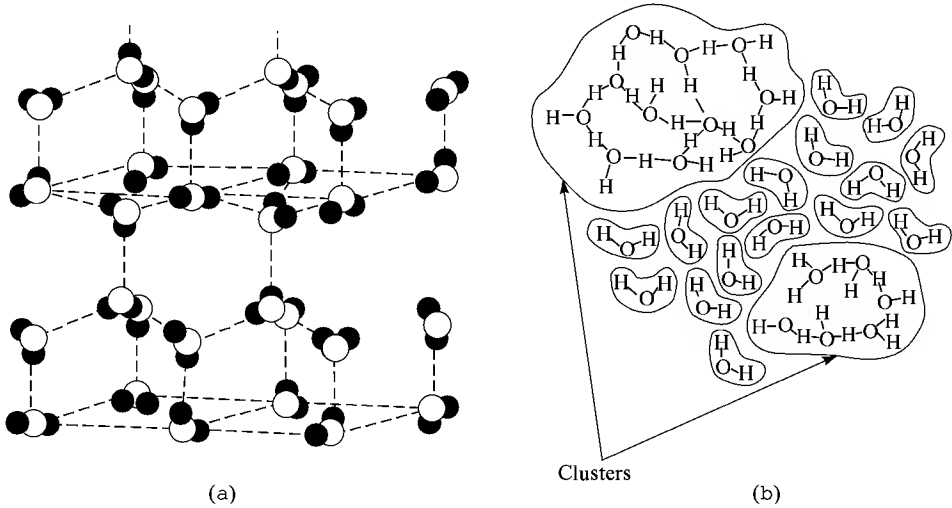


Fig. 5. (a) Hydrogen-bonded open tetrahedral structure of ice (4); (b) Frank-Wen flickering cluster model of liquid water (5).

Thermodynamic and physical properties of water vapor, liquid water, and ice I are given in Tables 3–5. The extremely high heat of vaporization, relatively low heat of fusion, and the unusual values of the other thermodynamic properties, including melting point, boiling point, and heat capacity, can be explained by the presence of hydrogen bonding (2,7).

Table 3. Selected Properties of Water Vapor^{a, b}

Property	Value
molecular weight	18.015
heat of formation, kJ mol ^c at 100°C	242.49
viscosity, mPa·s(cP) at 20°C	96 × 10 ⁻⁶
velocity of sound, m s at 100°C	405
diffusion coefficient, cm ² s at 100°C in air	0.380
specific volume, cm ³ g at 100°C	1729.6
specific heat, J (g·K) ^c at 100°C	2.078
thermal conductivity, W (cm·K) at 110°C	2.44 × 10 ⁻⁴

^a Ref. 6.
^b At 101.3 kPa (1 atm).
^c To convert J to cal, divide by 4.184.

Table 4. Selected Properties of Liquid Water^{a, b}

Property	Value
heat of formation, kJ mol ^c at 25°C	285.890

ionic dissociation constant, M^2 at 25°C	10^{-14}
heat of ionization, $\text{kJ}\cdot\text{mol}^{-1}$ at 25°C	55.71
apparent dipole moment, $\text{C}\cdot\text{m}^{\text{d}}$	6.24×10^{-30}
viscosity, $\text{mPa}\cdot\text{s}$ (cP) at 25°C	0.8949
velocity of sound, $\text{m}\cdot\text{s}$ at 25°C	1496.3
density, $\text{g}\cdot\text{cm}^{-3}$ at 25°C	0.9979751
at 0°C	0.99987
freezing point, °C	0.0
boiling point, °C	100.0
isothermal compressibility, over the range of 0.1–1 MPa, ^e nPa^{-1} at 25°C	0.45
specific heat at constant volume, $\text{J}\cdot(\text{g}\cdot\text{K})^{-1}$ at 25°C	4.17856
thermal conductivity, $\text{W}\cdot(\text{cm}\cdot\text{K})$ at 20°C	0.00598
temperature of maximum density, °C	3.98
dielectric constant, at 17°C and 60 MHz	81.0
electrical conductivity, $\text{S}\cdot\text{cm}$ at 25°C	$<10^{-8}$

^a Ref. 6.
^b At 101.3 kPa (1 atm).
^c To convert J to cal, divide by 4.184.
^d To convert $\text{C}\cdot\text{m}$ to D, divide by 3.336×10^{-30} .
^e To convert MPa to atm, divide by 0.101.

Table 5. Selected Properties of Ice ^{a, b}

Property	Value
heat of formation, $\text{kg}\cdot\text{mol}$ at 0°C	292.72
Young's modulus of elasticity, MPa^{c} at -10°C	967
density, $\text{g}\cdot\text{cm}^{-3}$ at 0°C	0.9168
coefficient of cubical thermal expansion, $\text{cm}^{-3}\cdot(\text{g}\cdot^{\circ}\text{C})$ at 0°C	120×10^{-6}
coefficient of linear thermal expansion, $^{\circ}\text{C}^{-1}$ at 0°C	52.7×10^{-6}
isothermal compressibility, nPa^{-1} at 0°C	0.12
specific heat, $\text{J}\cdot(\text{g}\cdot\text{K})$ at 0°C	2.06
thermal conductivity, $\text{W}\cdot(\text{m}\cdot\text{K})$	210
dielectric constant, at -1°C and 3 kHz	79

^a Ref. 6.
^b At 101.3 kPa (1 atm).
^c To convert MPa to psi, multiply by 145.

Solute Species

Dissolution of ionic and ionizable solutes in water is favored by ion–dipole bonds between ions and water. Figure 6 illustrates a hydrated sodium ion, Na^{+} (aq), eg, from dissolution of NaCl in water, surrounded by six water molecules in octahedral positions. The energy of the ion–dipole bonds depends on the size of the ion and its charge. Higher charge and smaller ionic radii favor the bonding. Further away from the central ion, the water molecules are structured through additional dipole–dipole interactions. In a similar way, the Cl^{-} (aq) ion interacts with the solvent water to form ion–dipole bonds, with the hydrogen side (local positive charges) of H_2O pointing toward the central ion.

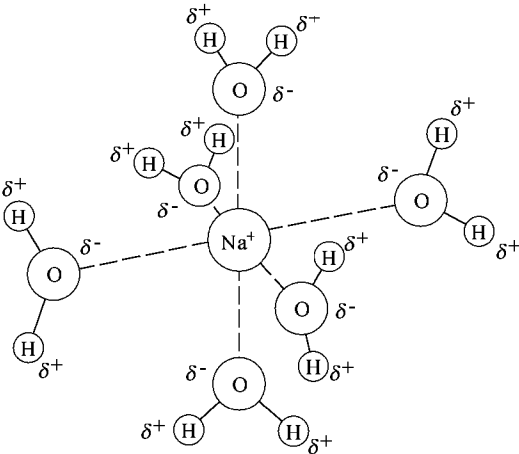


Fig. 6. Hydrated sodium ion, Na^{+} , in aqueous solution (4).The H_2O molecules form ion–dipole bonds to the central metal ion. The waters are in octahedral coordination to the sodium ion.

The dissolution of polar molecules in water is favored by dipole–dipole interactions. The solvation of the polar molecules stabilizes them in solution. Nonpolar molecules are soluble in water only with difficulty because the relatively high energy cost associated with disrupting and reforming the hydrogen-bonded water is unfavorable to the former occurring.

Hydrological Cycle and Water Reservoirs

In hydrological studies, the transfer of water between reservoirs is of primary interest. The magnitudes of the main reservoirs and fluxes (volume per time) are given in Figure 7. The oceans hold ca 76% of all the earth's water. Most of the remainder, ie, ca 21%, is contained in pores of sediments and in sedimentary rocks. A little more than 1% (or 73% of freshwater) is locked up in ice. The other freshwater reservoir of significant size is groundwater. Lakes, rivers, and the atmosphere hold a surprisingly small fraction of the earth's water.

The oceans are subdivided into surface (100–1000-m) ocean and deep ocean. The zone separating the warmer, surface water from the lower, cooler layer (oceanic thermocline) is characterized by a density gradient that prevents mixing.

The exchange of water between the various reservoirs includes ocean mixing, evaporation, precipitation, runoff, and percolation. The exchange of water between surface and deep ocean, ie, mixing by upwelling and downwelling, and the subsequent circulation of water masses are the most important water-transport processes. About 20 times more water is added to the surface ocean by mixing than by river runoff. The ocean–atmosphere–continents system may be regarded as a great heat engine; its water cycle is driven by the sun (see SOLAR ENERGY). Evaporation uses 23% of the energy of solar radiation. The oceans provide 86% and the continents 14% of the annual total atmospheric water. The average annual amount of worldwide precipitation is 0.7 m, ranging from 0 to 0.25 m for desert areas, to 0.25–0.40 m for grasslands or open woodland, 0.40–1.25 m for dry forests, and >1.125 m for wet forest. The mean residence time for a water molecule in the atmosphere is ca 10 d.

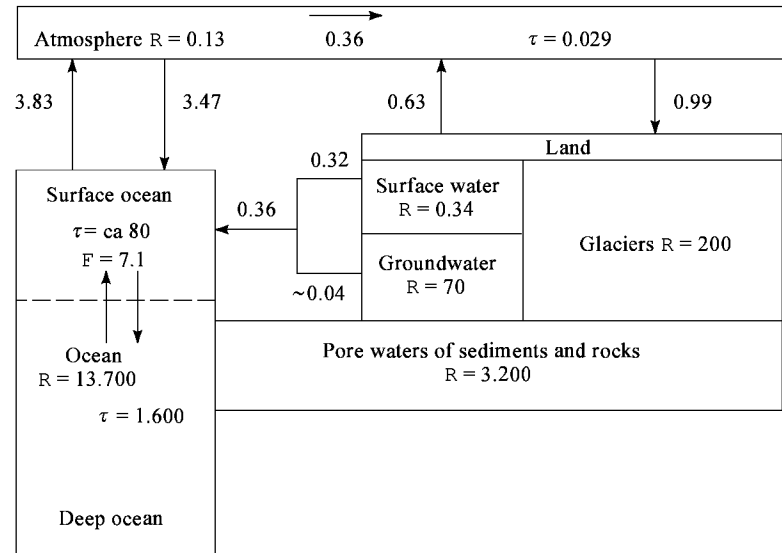


Fig. 7. The principal reservoirs in the hydrological cycle R , reservoirs in units of 10^{14} metric tons (10^5 km^3); F , fluxes in units of $10^5 \text{ km}^3 \text{ yr}^{-1}$; τ , residence time, yr; $R = F \times \text{volume input-output}$. Fluxes (flows) are approximate. For range of estimates, see Ref. 8.

Human Appropriation of Renewable Freshwater. Freshwater constitutes about 2.5% of the total volume of water on earth and two-thirds of this freshwater is locked in glaciers and ice caps.

Freshwater is now scarce in many regions of the world, resulting in severe ecological degradation, limits on agricultural and industrial production, threats to human health, and increased potential for international conflict.

As emphasized by Postel and co-workers (9), only freshwater flowing through the solar-powered hydrological cycle is renewable (Fig. 7). Nonreplenishable (fossil) groundwater can be tapped, but such extraction depletes reserves in much the same way as extractions from oil wells do. The terrestrial renewable freshwater supply, RFW_{land} , equals precipitation on land, P_{land} , which then subdivides into two major segments: evapotranspiration from the land, ET_{land} , and runoff to the sea, τ . Because groundwater and surface water are often hydraulically connected, soil infiltration and groundwater replenishment are included as part of this runoff component. Thus,

$$RFW_{\text{land}} = P_{\text{land}} - ET_{\text{land}} + r$$

Agriculture consumes by far the most of any use category to which the accessible runoff worldwide is applied (Table 6). Postel and co-workers estimate that human uses make up 26% of total terrestrial evapotranspiration and 54% of the runoff geographically and temporally accessible (9). Increased use of evapotranspiration will confer minimal benefits globally because most of the land suitable for rain-fed agriculture is already in production. New dam construction could increase accessible runoff by about 10% over the next 30 years; however, population increase during that period is projected to be more than 45%.

Table 6. Estimated Global Water Use and Consumption, by Sector, ca 1990^a

Sector	Use, km ³ yr	Consumption, km ³ yr
agriculture ^b	2880	1870
industry	975	90
municipalities	300	50
reservoir losses ^c	275	275
Subtotal	4430	2285
Instream flow needs	2350	0
Total	6780	2285
Total, percent of accessible runoff (12,500 km ³)	54	18

^a Ref. 9.
^b Assumes average applied water use of 12,000 m³ ha and consumption equal to ~65% of withdrawals.
^c Assumes evaporation loss equal to 5% of gross reservoir storage capacity.

Acquisition of Solutes and Circulation of Water with Rocks, Atmosphere, and Biota

Hydrogeochemical cycles couple atmosphere, land, and water. Natural waters acquire their chemical characteristics by dissolution and by chemical reactions

with solids, liquids, and gases with which they have come into contact during the various parts of the hydrological cycle. Waters vary in their chemical composition, but these variations are at least partially understandable if the environmental history of the water and the chemical reactions of the rock–water–atmosphere systems are considered. Dissolved mineral matter originates in the crustal materials of the earth; water disintegrates and dissolves mineral rocks by weathering. Gases and volatile substances participate in these processes. As a first approximation, seawater may be interpreted as the result of a gigantic acid–base titration, ie, acid of volcanoes vs bases of rocks (oxides, carbonates, silicates) (10). The composition of freshwater may similarly be represented as resulting from the interaction of the CO₂ of the atmosphere with mineral rocks.

Figure 8, inspired by Siever (11), gives a simplified picture of some of the more important global biogeochemical transformations. A useful analogy would be that these are linked by a multitude of valves and switches that in effect control the system. Of particular interest are the gas regulators of the CO₂ and O₂ tanks. Both O₂ and CO₂ regulators are governed by photosynthesis and by weathering (11).

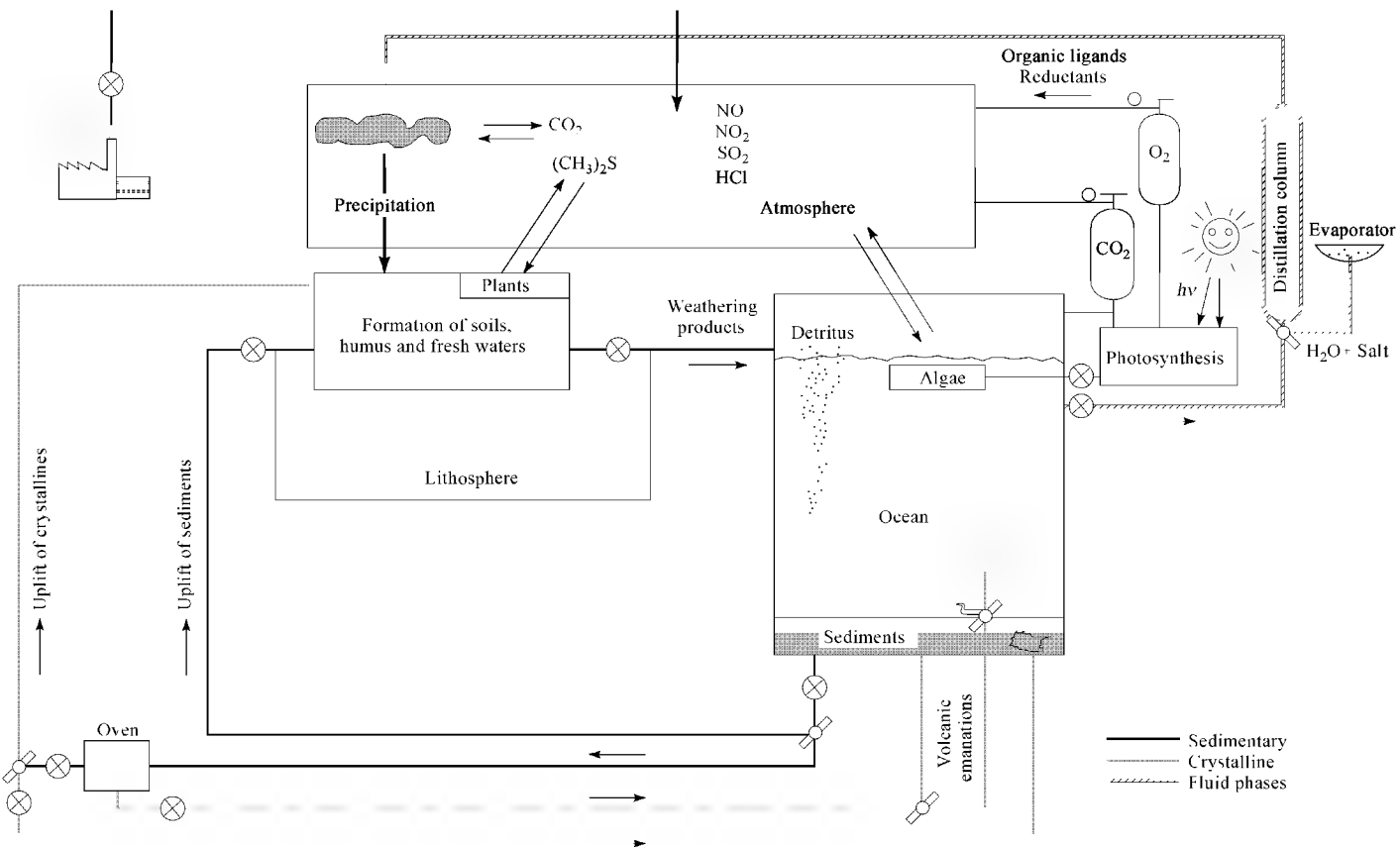


Fig. 8. Steady-state model for the earth's surface geochemical system. The interaction of water with rocks in the presence of photosynthesized organic matter continuously produces reactive material of high surface area. This process provides nutrient supply to the biosphere and, along with biota, forms the array of small particles (soils). Weathering imparts solutes to the water, and erosion brings particles into surface waters and oceans.

The classical geochemical material balance (12) assumes that the H⁺ balance and the electron balance (oxidation state) in our environment have been established globally by the interaction of primary (igneous) rocks with volatile substances (Table 7).

Table 7. Geochemical Material Balance

Igneous rocks	Volatile substances	Seawater	Atmosphere	Sediments (1)
silicates	CO ₂	pH 8	p _{O₂} 0.2 atm	carbonates
carbonates	H ₂ O		p _{CO₂} 0.0003 atm	silicates
oxides	SO ₂			
	HCl			

In an oversimplified way, it may be stated that acids of the volcanoes have reacted with the bases of the rocks; the compositions of the ocean (which is at the first end point (pH = 8) of the titration of a strong acid with a carbonate) and the atmosphere (which with its p_{CO₂} = 10^{-3.5} atm atm is nearly in equilibrium with the ocean) reflect the proton balance of reaction 1. Oxidation and reduction are accompanied by proton release and proton consumption, respectively. In order to maintain charge balance, the production of electrons, e⁻, must eventually be balanced by the production of H⁺. The redox potential of the steady-state system is given by the partial pressure of oxygen (0.2 atm). Furthermore, the dissolution of rocks and the precipitation of minerals are accompanied by H⁺ consumption and H⁺ release, respectively.

Atmosphere–Water Interaction. Although water is a very minor component of the atmosphere, less than 10⁻⁶ vol % of the atmosphere consisting of water, many important reactions occur in the water droplets of cloud, fog, and rain. The atmosphere is an oxic environment; in its water phase, gigantic quantities of reductants, such as organic substances, Fe(II), SO₂, CH₃SCH₃ (dimethyl sulfide), and nitrogen oxides, are oxidized by oxidants such as oxygen, OH· radicals, H₂O₂, and Fe(III).

The atmosphere is an important conveyor belt for many pollutants. The atmosphere reacts most sensitively to anthropogenic disturbance because proportionally it represents a much smaller reservoir than land and water; furthermore, the residence times of many constituents of the atmosphere are smaller than those that occur in the other exchange reservoirs. Water and atmosphere are interdependent systems. Many pollutants, especially precursors of acids and photooxidants, originate directly or indirectly from the combustion of fossil fuels. Hydrocarbons, carbon monoxide, and nitrogen oxides released by thermal power plants and, above all by automobile engines, can produce, under the influence of sunlight, ozone and other photooxidants.

Acid Rain. Figure 9 shows the various processes that involve atmospheric pollutants and natural components in the atmosphere. The following reactions are of particular importance in the formation of acid precipitation: oxidative reactions, either in the gaseous phase or in the aqueous phase, leading to the formation of oxides of C, S, and N (CO₂; SO₂, SO₃, H₂SO₄; NO, NO₂, HNO₂, HNO₃); absorption of gases into water (cloud droplets, falling

raindrops, or fog) and interaction of the resulting acids ($\text{SO}_2 \cdot \text{H}_2\text{O}$, H_2SO_4 , HNO_3) with ammonia (NH_3) and the carbonates of airborne dust; and the scavenging and partial dissolution of aerosols into water. In this case, aerosols are produced from the interaction of vapors and airborne (maritime and dust) particles; they often contain $(\text{NH}_4)_2\text{SO}_4$ and NH_4NO_3 .

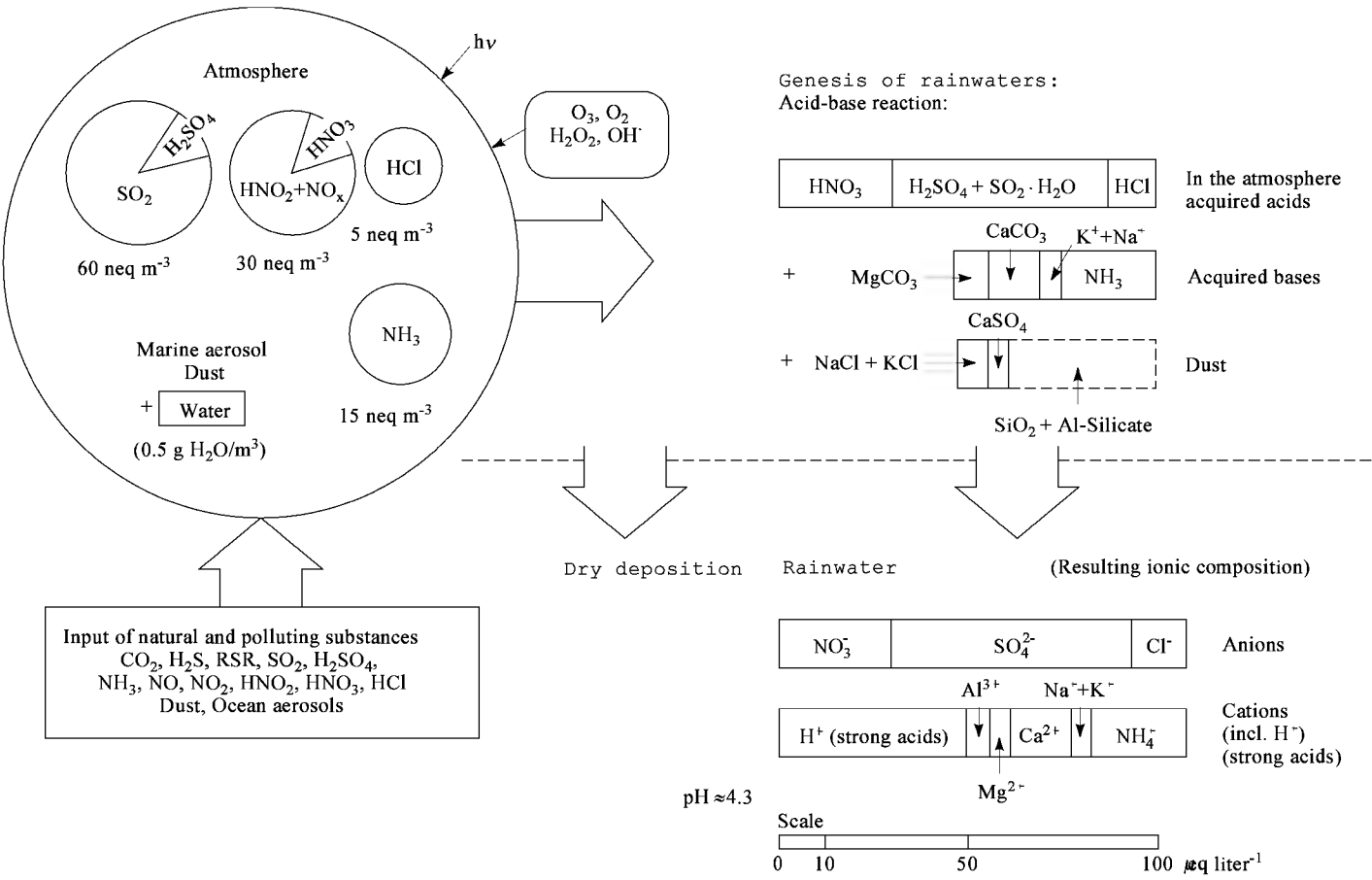


Fig. 9. Genesis of acid rain (13). From the oxidation of C, S, and N during the combustion of fossil fuels, there is a buildup in the atmosphere (gas phase, aerosol particles, raindrops, snowflakes, and fog) of CO_2 and the oxides of S and N, which leads to acid–base interaction. The importance of absorption of gases into the various phases of gas, aerosol, and atmospheric water depends on a number of factors. The genesis of acid rain is shown on the upper right as an acid–base titration. The data given are representative of the environment in the vicinity of Zürich, Switzerland.

Deposition. The products of the various chemical and physical reactions in the atmosphere are eventually returned to the earth's surface. Usually, a useful distinction is made here between wet and dry deposition. Wet deposition, ie, rainout and washout, includes the flux of all those components that are carried to the earth's surface by rain or snow, that is, those dissolved and particulate substances contained in rain or snow. Dry deposition is the flux of particles and gases, especially SO_2 , HNO_3 , and NH_3 , to the receptor surface during the absence of rain or snow. Deposition can also occur through fog, aerosols and droplets which can be deposited on trees, plants, or the ground. With forests, approximately half of the deposition of SO_4^{2-} , NH_4^+ , and H^+ occurs as dry deposition.

The rainwater shown in Figure 9 contains an excess of strong acids, most of which originate from the oxidation of sulfur during fossil fuel combustion and from the fixation of atmospheric nitrogen to NO and NO_2 , eg, during combustion of gasoline by motor vehicles. It should also be mentioned that there are natural sources of acidity, resulting from volcanic activity, from H_2S from anaerobic sediments, and from dimethyl sulfide and carbonyl sulfide that originate in the ocean. HCl results from the combustion and decomposition of organochlorine compounds such as poly(vinyl chloride). Bases originate in the atmosphere as the carbonate of wind-blown dust and from NH_3 , generally of natural origin. The NH_3 comes from NH_4^+ and from the decomposition of urea in soil and agricultural environments.

The reaction rates for oxidation of atmospheric SO_2 (0.05–0.5 d⁻¹) yield a sulfur residence time of several days, at most; this corresponds to a transport distance of several hundred to 1000 km. The formation of HNO_3 by oxidation is more rapid and, compared with H_2SO_4 , results in a shorter travel distance from the emission source. H_2SO_4 can also react with NH_3 to form NH_4HSO_4 or $(\text{NH}_4)_2\text{SO}_4$ aerosols. In addition the NH_4NO_3 aerosols are in equilibrium with NH_3 (g) and HNO_3 (g).

In addition to the acid–base components shown in Figure 9, various organic acids are often found. Many of these acids are by-products of the atmospheric oxidation of organic matter released into the atmosphere. Of special interest are formic, acetic, oxalic, and benzoic acids, which have been found in rainwater in concentrations occasionally exceeding a few micromoles per liter.

Groundwater. The uppermost part of the earth's rocks constitutes a porous medium in which water is stored and through which it moves. Up to a certain level, these rocks are saturated with water that is free to flow laterally under the influence of gravity. Subsurface water in this saturated zone is groundwater, and the uppermost part of the zone is the water table. The chemistry of the groundwater is influenced by the composition of the aquifer and by the chemical and biological events occurring in the infiltration.

A general relationship between the composition of the water and that of the solid minerals with which the water has come into contact during infiltration and in the aquifer can be expected. Biological activity, especially in the organic layer above the mineral part, has a pronounced effect on the acquisition of solutes. Because of microbial respiration, the CO_2 pressure is increased. CO_2 pressure tends to increase the alkalinity and the concentration of Ca^{2+} and other solutes.

Interactions with Rocks; Chemical Weathering

Chemical weathering is one of the major processes controlling the global hydrogeochemical cycle of elements (14–16). In this cycle, water operates both as a reactant and as a transporting agent of dissolved and particulate components from land to sea. The atmosphere provides a reservoir for carbon dioxide and for oxidants required in the weathering reactions. The biota assists the weathering processes by providing organic ligands and acids and by supplying locally, upon decomposition, increased CO_2 concentrations.

During chemical weathering, rocks and primary minerals become transformed to solutes and soils, and eventually to sediments and sedimentary

rocks. Weathering reactions consume protons and produce alkalinity.

Under natural conditions the rates of dissolution of most minerals are too slow to depend on mass transfer of the reactants or products in the aqueous phase. This restricts the case to one either of weathering reactions where the rate-controlling mechanism is the mass transfer of reactants and products in the solid phase, or of reactions controlled by a surface process and the related detachment process of reactants.

Calcareous minerals and evaporite minerals (halides, gypsum) are very soluble and dissolve rapidly and, in general, congruently, ie, yielding upon dissolution the same stoichiometric proportions in the solution as the proportions in the dissolving mineral and without forming new solid phases (Fig. 10). Their contribution to the total dissolved load in rivers can be estimated by considering the mean composition of river water and the relative importance of various rocks to weathering. Estimates (18) indicate that evaporites and carbonates contribute approximately 17% and 38%, respectively, of the total dissolved load in the world's rivers. The remaining 45% is the result of the weathering of silicates, underlining the significant role of these minerals in the overall chemical denudation of the earth's surface.

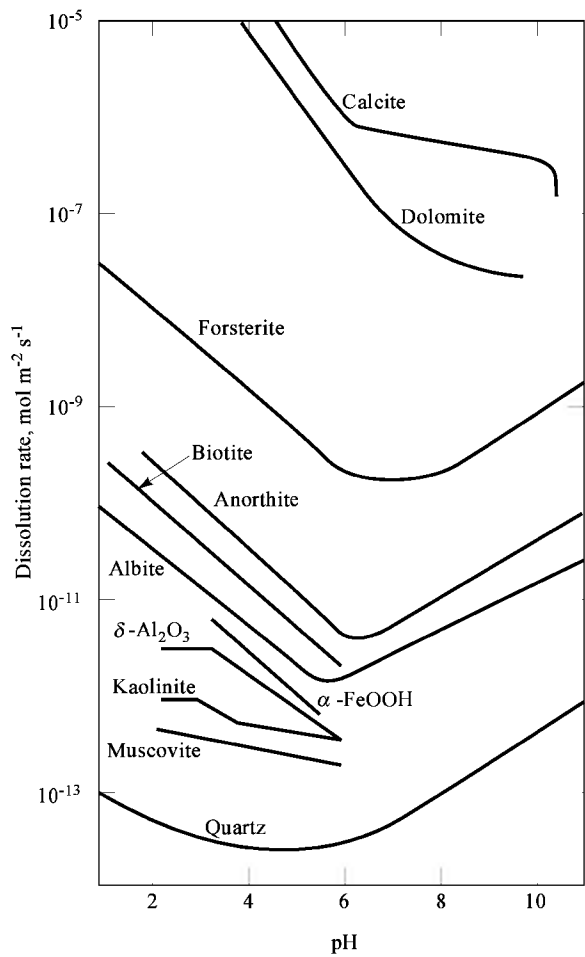
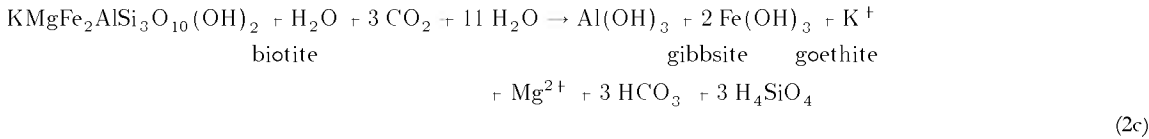
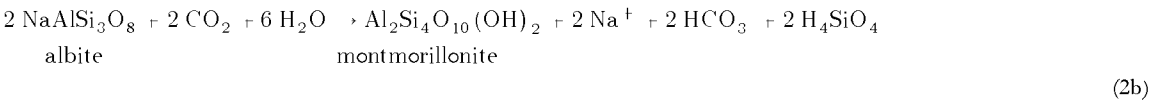
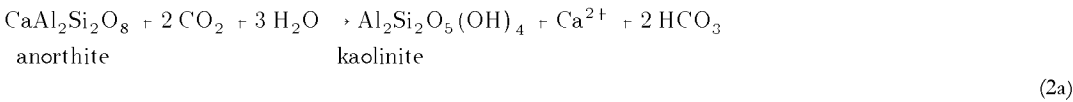


Fig. 10. Dissolution rate of different minerals as a function of pH at 25°C (7,17).

There are no unequivocal weathering reactions for the silicate minerals. Depending on the nature of parent rocks and hydraulic regimes, various secondary minerals like gibbsite, kaolinite, smectites, and illites are formed as reaction products. Some important dissolution processes of silicates are given, for example, by the following reactions (19).



In all cases, water and carbonic acid, the latter of which is the source of protons, are the main reactants. The net result of the reaction is the release of cations (Ca²⁺, Mg²⁺, K⁺, Na⁺) and the production of alkalinity via HCO₃⁻. When ferrous iron is present in the lattice, as in biotite, oxygen consumption may become an important factor affecting the weathering mechanism and the rate of dissolution.

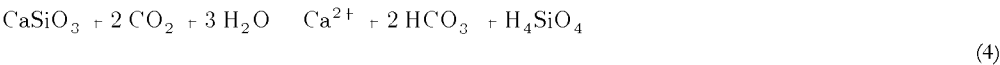
These reactions, however, are complex and generally proceed through a series of reaction steps. The rate of weathering of silicates may vary considerably, depending on the arrangement of the silicon tetrahedra in the mineral and on the nature of the cations.

The Role of Weathering in Geochemical Processes in Oceanic and Global Systems. As indicated in the stoichiometric equations given above, the rate of chemical weathering is important in determining the rate of CO₂ consumption. Furthermore, the global weathering rate is most likely influenced by temperature and is proportional to total continental land area and the extent of its coverage with vegetation; the latter dependence results from CO₂ production in soils, which is a consequence of plant respiration and the decay of organic matter as well as the release of complex-forming substances (ligands), eg, dicarboxylic acids, hydroxycarboxylic acids, and phenols, ie, anions that form soluble complexes with cations that originate from the lattice or form surface complexes with the surface of oxide minerals. Regionally, the extent of weathering is influenced by acid rain.

It has been shown (20,21) that silicate weathering is more important than carbonate mineral weathering as a longterm control on atmospheric CO₂. The HCO₃⁻ and Ca²⁺ ions produced by weathering of CaCO₃ precipitate in the



ocean as CaCO₃ through incorporation by marine organisms. The CO₂ consumed in carbonate weathering is released again upon formation of CaCO₃ in the ocean in a reversal reaction 3. Thus, globally, carbonate weathering results in no net loss of atmospheric CO₂. The weathering of calcium silicates, eg, equation 2a or, in a simplified way, equation 4, also produces Ca²⁺ and HCO₃⁻, which form



CaCO₃ in the sea, but only half of the CO₂ consumed in the weathering is released and returned to the atmosphere upon CaCO₃ formation. Thus, silicate weathering results in a net loss of atmospheric CO₂.

The dissolution processes are influenced by the environment, including pH, complex formation, and changes in the oxidation state (Fig. 10). Naturally occurring organic matter, such as water-soluble anions of low molecular weight or low molecular weight organic acids, may form chelate complexes, whereas natural redox processes may control the pH. For example, the oxidation of pyrite yields ferric oxide and acid which may react with the rocks; on the other hand, hematite, Fe₂O₃, may, upon reduction, produce large quantities of OH⁻. Physical and biological weathering may accelerate the relatively slow chemical weathering process by further comminuting inorganic material, thereby increasing the specific surface area susceptible to weathering.

Biochemical Cycle and Oxidation–Reduction Processes

The ecological system may be considered as a unit of the environment that contains a biological organization made up of all the organisms interacting reciprocally with the chemical and physical environment. The maintenance of life in aquatic ecosystems results directly or indirectly from the steady impact of solar energy. Photosynthesis is carried out mostly by algae and water plants; it may be conceived as a disproportionation of water into an oxygen reservoir and hydrogen, which forms high energy bonds with C, N, S, and P compounds that are incorporated as organic matter in the biomass. Various organisms catalytically decompose the unstable products of photosynthesis through energy-yielding electron-transfer reactions, eg, reduction–oxidation processes and respiration. These organisms use this source of energy for their metabolic needs (7,16).

The composition of natural waters is strongly influenced by the growth, distribution, and decay of algae and other organisms. Organisms regulate the oceanic and lacustrine composition and its variation with depth. Dissolved constituents are taken up by organisms. Their remains sink under the influence of gravity and are gradually destroyed by oxidation. The superposition of this particular cycle upon the ordinary mixing cycle in the ocean or in lakes accounts for the variation in depth and distribution of chemical properties.

Chemical Composition

The concentrations of the principal elements in marine and freshwaters are given in Figure 11. The concentrations of biologically regulated components, ie, C, N, P, and Si, vary with depth and are markedly influenced by the growth, distribution, and decay of phytoplankton and other organisms. The concentrations of other constituents, especially the salts, ie, Cl⁻, SO₄²⁻, Mg²⁺, and Na⁺, are remarkably constant and are different from those in fresh waters. For most elements, a mass balance appears to exist between input into the sea, mostly by rivers, and its removal, mostly by sedimentation. Thus, the oceans are often assumed to be at steady state. The mean residence time, τ, (yr) of an element or constituent is defined as average time a given element spends in the ocean before being removed by sedimentation and other processes, where *M* is the concentration of

$$\tau = \frac{\text{total mass}}{\text{input or output}} = \frac{M}{Q}$$

a particular constituent in seawater multiplied by the volume of the ocean water, and *Q* is the product of the concentration of the particular element in average river water and the annual flux of river water to the ocean or the quantity of this element being removed from the water by sedimentation. Some sea salts, especially NaCl, are recycled through the atmosphere as aerosols from droplets of sea spray to the land masses, where they are washed out by rain. Chloride and sodium originating from the atmosphere, for example, contribute about one-third of the annual river flux of these elements. The other main constituents are derived principally from the weathering of rocks, and the fraction for atmospherically cycled salts is small.

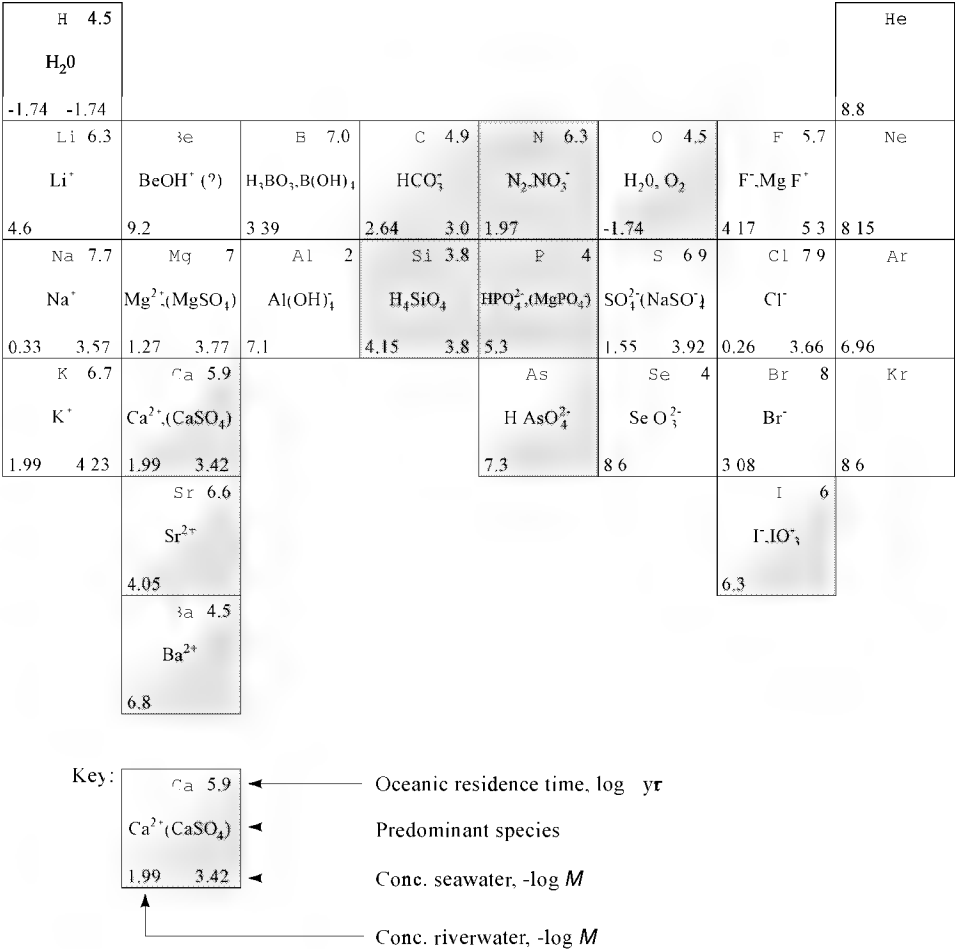


Fig. 11. Elements in natural waters, their form of occurrence, and concentration. Elements whose distribution is significantly affected by biota are shaded; P, N, and Si (fully shaded) are often depleted in surface waters. Species in parentheses are the main ion pairs in seawater. Data are uncorrected for atmospheric recycling of marine salts.

The chemistry of natural waters is covered in References 7, 22, and 23. The aquatic environmental chemistry of soils is reviewed in Reference 24. The chemistry of the mineral–water interface and of aquatic environmental particles and colloids is reviewed in References 25 and 26. References 16 and 27 review the role of the hydrosphere in the biogeochemistry of global change.

The concept of residence time can also be applied to lakes where the flow through the outlet has to be considered. In lakes it is often convenient to define a relative residence time, ie, a residence time relative to that of water.

The flow of hydrothermal solutions into the oceans from hydrothermal vents, ie, springs coming from the sea floor in areas of active volcanism, and the chemical reactions occurring there by high temperature alteration of basalts are of significance in the mass balance of Mg²⁺ and K⁺. Furthermore, during the cycling of seawater through the earth's crust along the mid-ocean ridge system, geothermal energy is transferred into chemical energy in the form of reduced inorganic compounds. These compounds are derived from the reaction of seawater with crustal rocks at high temperatures and are emitted from warm (<25°C) and hot (~350°C) submarine vents at depths of 2000–3000 m. Chemolithotrophic bacteria use these reduced chemical species as sources of energy for the reduction of carbon dioxide (assimilation) to organic carbon (28).

The observed oceanic residence times range from 100 yr for aluminum to ca 10⁸ yr for chlorine (Figs. 5 and 6; Table 8), reflecting the great variation in geochemical reactivity; unreactive elements simply accumulate in the ocean and have large τ values. Figure 12 shows the strong correlation between the mean oceanic residence time of an element and the distribution of that element between seawater and crystal rocks (30).

Table 8. Major Composition of Seawater

Constituent	Seawater at $S = 35$, ^a	(g kg ⁻¹) : Chlorinity ^b	(mol kg ⁻¹) : Chlorinity	Residence time in oceans, log τ , years ^c
Na ⁺	10.77	0.556	0.0242	7.7
Mg ²⁺	1.29	0.068	0.0027	7.0
Ca ²⁺	0.4121	0.02125	0.000530	5.9
K ⁺	0.399	0.0206	0.000527	6.7
Sr ²⁺	0.0079	0.00041	0.0000047	6.6
Cl	19.354	0.9989	0.8282	7.9
SO ₄ ²⁻	2.712	0.1400	0.0146	6.9
HCO ₃ ⁻	0.1424	0.00735	0.00012	4.9
Br	0.0673	0.00348	0.000044	8
F	0.0013	0.000067	0.0000035	5.7
B	0.0045	0.000232	0.0000213	7.0
$\Sigma = 35$		$\Sigma = 1.82$	$\Sigma = 0.058$	

^a Salinity, $S(\text{‰})$, is defined as the weight in grams of the dissolved inorganic matter in 1 kg of seawater after all Br⁻ and I⁻ have been replaced by the equivalent quantity of Cl⁻ and HCO₃⁻ and CO₃²⁻ are converted to oxide. In over 97% of the seawater in the world, the salinity S is between 33‰ and 37‰. The total number of grams of major constituents (sea salt), g_T , for 1 kg of solution is related to the chlorinity by $g_T = 1.81578 \text{ Cl } (\text{‰})$. Salinity S (‰), is defined as $S(\text{‰}) = 1.80655 \text{ Cl } (\text{‰})$; thus $g_T = 1.00511 S$ (‰).

^b The chlorinity, Cl (‰), is determined by the titration of seawater with AgNO₃. It was defined as the chlorine equivalent of the total halide concentration in g kg⁻¹ seawater; it is now defined as the mass in grams of Ag necessary to precipitate the halogens (Cl⁻ and Br⁻) in 328.5233 g of seawater. It has been adequately demonstrated that the relative composition of the major (greater than 1 m kg⁻¹ seawater) components of seawater is nearly constant. By measuring one constituent of seawater, the composition of other components can be characterized. The constituent normally selected is the chlorinity, Cl

(‰).

^c Residence times were computed by $\tau = M/Q$, where M for a particular constituent is equal to its concentration in seawater times the mass of the oceans, and Q is equal to the concentration of the constituent in average river water times the annual flux of river water to the ocean.

^d The results given for HCO_3^- are actually values of the carbonate alkalinity expressed as though it were all HCO_3^- .

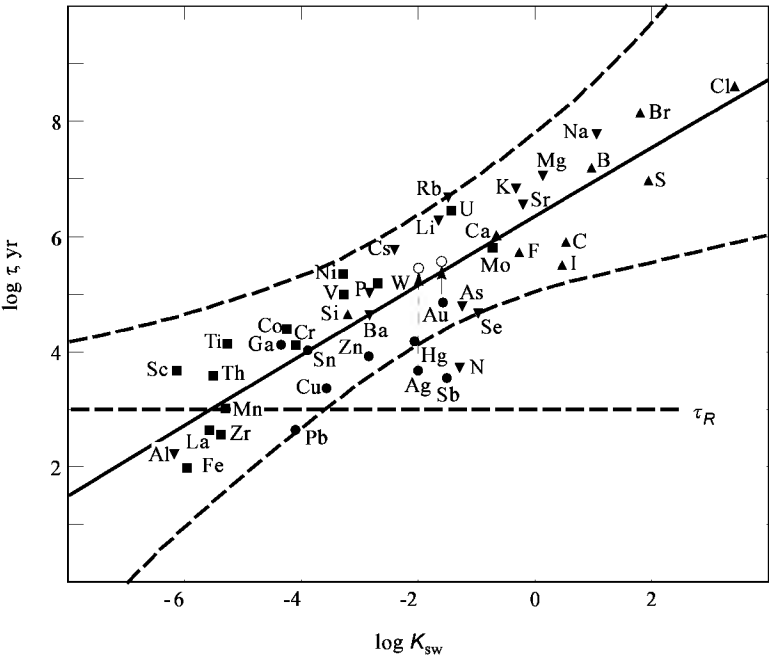


Fig. 12. The relationship between the mean oceanic residence time, τ , yr, and the seawater–crustal rock partition ratio, K_{sw} , of the elements; adapted from Ref. 29. ▽, Pretransition metals; ■, transition metals; ●, B-metals; ▲, nonmetals. Open symbols indicate τ -values estimated from sedimentation rates. The solid line indicates the linear regression fit, and the dashed curves show the Working-Hotelling confidence band at the 0.1% significance level. The horizontal broken line indicates the time required for one stirring revolution of the ocean, τ_R .

The transition between rivers and oceans occurs in estuaries, which are semienclosed coastal bodies of water within which seawater is diluted with freshwater from land drainage. Several types of water circulation may occur (vertically mixed or stratified), depending on topography, tidal currents, and freshwater discharge. Estuaries are of great environmental concern. Many marine organisms require the estuarine environment for part of their life cycle; thus, pollution in an estuary may have long-range ecological consequences.

Estuaries with their salinity gradients and tidal movements represent gigantic natural coagulation tanks. In coagulation, colloidal particles agglomerate and are subsequently removed by sedimentation. Clay minerals, organic matter, colloidal hydroxides of iron(III), and heavy metals tend to coagulate and accumulate in the estuary. The trapping of organic matter may impart anaerobic conditions in the lower part of estuarine water.

Chemical Variety. The term species refers to the actual form in which a molecule or ion is present in solution. For example, a metal ion may occur in natural waters, as a free metal ion, ie, an aquo complex $\text{Me}(\text{H}_2\text{O})_n^{z+}$, an inorganic or organic complex, and it may be present in dissolved or particulate, often colloidal, form. It is difficult to distinguish between the dissolved and particulate forms. The definition by pore size (0.45 μm) of membrane filter is considered questionable, and what are thought to be dissolved species may often include some or most of the colloidal forms of an element. The inorganic fraction of suspended particulate matter consists mainly of minerals formed by weathering of terrestrial rocks. The organic particulate matter is primarily composed of living organisms, their decay, and metabolic products. Inorganic trace elements may be sorbed or bonded onto them.

Dissolved inorganic species present in seawater are principally electrolytes (Table 8), uncharged species, and dissolved gases. The species patterns for cations in model seawater (pH 8.2) and freshwater have been derived from calculations based on stability constants of complexes (7,17,31). The principal species of the more important elements occurring in natural waters are given in Figure 11. The concentration and distribution of dissolved trace metals in surface marine waters has been reviewed (23,38). A few typical concentrations of dissolved heavy metals in natural waters are given in Table 9. The characterization of the carbon dioxide system in marine waters is given in Reference 39.

Table 9. Representative Concentration of Some Dissolved Heavy Metals in Natural Waters

Area or source	Metal, nM				Reference
	Cu	Zn	Cd	Pb	
East coast U.S. rivers	17	13	0.095	0.11	32
Mississippi River	23	3	23		33
Amazon River	24	0.3–0.8	0.06		33
Lake Constance	5–20	15–60	0.05–0.1	0.2–0.5	34
Lake Michigan	10	9	0.17	0.25	35
Lago Cristallina ^a	5	30	0.5	3	36
Pacific Ocean	0.5–5	0.1–10	0.01–1	0.005–0.08	37
rainwater ^b	10–300	80–900	0.4–7	10–200	36

^a Alpine lake at 2200-m altitude in the southern Swiss Alps. Its dissolved Al(III) concentration at pH6 = 600 nM.

^b As measured near Zürich, Switzerland.

Metals and metalloids that form alkyl compounds, eg, methylmercury and methylarsenic acid, tributyltin, deserve special concern because these compounds are volatile and accumulate in cells; they are poisonous to the central nervous system of higher organisms. Because methylmercury or other metal alkyls may be produced at a rate faster than it is degraded by other organisms, it may accumulate in higher organisms such as fish. Hg species are also reduced to elementary Hg^0 which is soluble in water but lost by volatilization to the atmosphere (40).

Pollution

The products of human activities find their way into the environment and disturb ecosystems. Pollution has altered the surroundings to the detriment of humanity. In the last several decades, the pollutional load has increased, and its character has changed (see WATER–POLLUTION).

The balance between photosynthesis and respiration in a receiving water, ie, a lake, estuary, or coastal basin, may become perturbed by pollution with inorganic plant (algal) nutrients, eg, phosphorus and nitrogen compounds (eutrophication). Excessive production of algae in surface waters is followed, subsequent to the settling of the biota and its debris into the deeper waters, by the respiration (oxidative mineralization) of this organic material by oxygen, creating an oxygen deficit or even anaerobic conditions in the deeper waters. Many trace constituents, eg, mercury, pesticides, phenols, are harmful to specific aquatic life and other organisms.

The fate of a pollutant in an aquatic system may be expressed as follows:

$$\text{accumulation} = \text{input} - (\text{interphase} + \text{intrapphase mass transfer}) - (\text{chemical} + \text{biological reaction}) - \text{output}$$

By evaluating the strength of anthropogenic and natural emission sources, identifying the relevant pathways, and characterizing the various processes (dilution, dispersion, transport, partitioning, adsorption, volatilization, chemical and biological transformation, and biodegradation) that govern the behavior of chemicals in the environment, present and future fluxes can be predicted as well as the distribution, fate, and residual activities of pollutants and their effect on ecological systems and on human health (41).

Pollutant Distribution. Of particular importance for the aquatic ecosystem is the distribution of volatile substances, eg, gases and volatile organic compounds, between the atmosphere and water, and the sorption of compounds at solid surfaces, eg, settling suspended matter, biological particles, sediments, and soils (41,42).

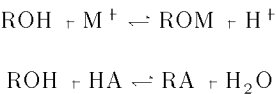
Hydrophobic substances are soluble in nonpolar solvents, whereas their solubility in water is very limited. Many of these substances are also soluble in fats and lipids and are also called lipophile compounds. Such substances have a tendency to avoid contact with water and to associate with a nonpolar, nonaqueous environment, such as a surface, eg, an organic particle, a particle containing organic material, or the lipid-containing biomass of an organism.

The solubility of hydrophobic substances in, or their absorbability on suspended particles, on sediments, on biota, or on soil particles can be related to the solubility of these substances in organic solvents. The solvent *n*-octanol, CH₃(CH₂)₇OH, is a kind of surrogate for many kinds of environmental and physiological organic substances and has become a reference phase for organic phase water partitioning of organic solutes.

The accumulation of xenobiotic substances in organisms and the food chain (bioaccumulation) is important in the assessment of the harmfulness of a substance in the environment. The accumulation of organic substances in organisms often occurs in accordance with their lipophilicity. The partitioning into octanol often serves as a measure of lipophilicity and as a predictive parameter for bioaccumulation in the food chain. One also speaks of biomagnification to describe progressive accumulation of xenobiotic substances in the food chain. Partitioning models are based on the assumption that chemicals of concern partition in a more or less passive fashion between the aquatic environment and the organism living in that environment.

Under equilibrium or near-equilibrium conditions, the distribution of volatile species between gas and water phases can be described in terms of Henry's law. The rate of transfer of a compound across the water–gas phase boundary can be characterized by a mass-transfer coefficient and the activity gradient at the air–water interface. In addition, these substance-specific coefficients depend on the turbulence, interfacial area, and other conditions of the aquatic systems. They may be related to the exchange constant of oxygen as a reference substance for a system-independent parameter; reaeration coefficients are often known for individual rivers and lakes.

Adsorption of Metal Ions and Ligands. The solid–solution interface is of greatest importance in regulating the concentration of aquatic solutes and pollutants. Suspended inorganic and organic particles and biomass, sediments, soils, and minerals, eg, in aquifers and infiltration systems, act as adsorbents. The reactions occurring at interfaces can be described with the help of surface-chemical theories (surface complex formation) (25). The adsorption of polar substances, eg, metal cations, M, anions, A, and weak acids, HA, on hydrous oxide, clay, or organically coated surfaces may be described in terms of surface-coordination reactions:



With the help of equilibrium constants, the extent of adsorption can be predicted as a function of pH and solution variables (7,25,43). Based on this model, the partitioning of metal ions and of ligands (organic and inorganic anions between water and pelagic clays and suspended particles) can be explained.

BIBLIOGRAPHY

"Water Properties" in *ECT* 2nd ed., Vol. 21, pp. 668–688, by R. A. Horne, Woods Hole Oceanographic Institution; "Properties" under "Water" in *ECT* 3rd ed., Vol. 24, pp. 276–294, by C. Matter-Möller and W. Stumm, Swiss Federal Institute of Technology.

1. D. Eisenberg and W. Kauzmann, *The Structure and Properties of Water*, Oxford University Press, London, 1969.
2. E. K. Berner and R. A. Berner, *The Global Water Cycle: Geochemistry and Environment*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1987.
3. R. A. Horne, *Marine Chemistry*, Wiley-Interscience, New York, 1969.
4. H. B. Gray, *Chemical Bonds*, Benjamin Cummings, Menlo Park, Calif., 1973.
5. G. Nemethy and H. A. Scheraga, *J. Chem. Phys.* **36**, 3382 (1962).
6. R. A. Horne, ed., *Water and Aqueous Solutions: Structure, Thermodynamics, and Transport Processes*, Wiley-Interscience, New York, 1972.
7. W. Stumm and J. J. Morgan, *Aquatic Chemistry*, 3rd ed., Wiley-Interscience, New York, 1996.
8. P. H. Gleick, ed., *Water in Crisis, A Guide to the World's Fresh Water Resources*, Oxford University Press, Oxford, U.K., 1993.
9. S. L. Postel, G. C. Daily, and P. R. Ehrlich, *Science* **271**, 785 (1996).
10. L. G. Sillin in M. Sears, ed., *Oceanography*, American Association for the Advancement of Science (AAAS), Washington, D.C., 1961, pp. 549–581.
11. R. Siever, *Sedimentology* **11**, 5 (1968).
12. V. M. Goldschmidt, *Fortschr. Mineral. Krist. Petrol.* **17**, 112 (1933).
13. J. L. Schnoor and W. Stumm, in W. Stumm, ed., *Chemical Processes in Lakes*, Wiley-Interscience, New York, 1985, pp. 311–338.
14. J. I. Drever, *The Geochemistry of Natural Waters*, 2nd ed., Prentice-Hall, Inc., Englewood Cliffs, N.J., 1988.
15. R. Wollast in W. Stumm, ed., *Aquatic Chemical Kinetics, Reaction Rates of Processes in Natural Water*, Wiley-Interscience, New York, 1990, pp. 431–445.
16. E. K. Berner and R. A. Berner, *Global Environment: Water, Air, and Geochemical Cycles*, Prentice Hall, Inc., Englewood Cliffs, N.J., 1996.
17. L. Sigg and W. Stumm, *Aquatische Chemie, Eine Einführung in die Chemie wässriger Lösungen und natürlicher Gewässer* 3rd ed., Teubner, Stuttgart, Germany, 1994.
18. R. Wollast and F. T. Mackenzie, in S. R. Aston, ed., *Silicon Geochemistry and Biogeochemistry*, Academic Press, New York, 1983, pp. 39–76.
19. W. Stumm and R. Wollast, *Revs. Geophys.* **28**(1), 53–69 (1990).
20. R. A. Berner, A. C. Lasaga, and R. M. Garrels, *Amer. J. Sci.* **283**, 641–683 (1983).
21. R. A. Berner and A. C. Lasaga, *Scientif. Amer.* **260**, 74–81 (1989).
22. J. D. Buffle, *Complexation Reactions in Aquatic Systems: An Analytical Approach*, Ellis Horwood, Chichester, U.K., 1988.

23. F. M. M. Morel and J. G. Hering, *Principles and Applications of Aquatic Chemistry*, Wiley-Interscience, New York, 1993, pp. 374.

24. G. Sposito, *Chimia* **43**, 169–176 (1989).

25. W. Stumm, *Chemistry of the Solid–Water Interface: Processes at the Mineral–Water and Particle–Water Interface*, Wiley-Interscience, New York, 1992.

26. J. D. Buffle and H. P. van Leeuwen, eds., *Environmental Particles*, Vol. 1, Lewis, Boca Raton, Fla., 1992.

27. W. H. Schlesinger, *Biogeochemistry, An Analysis of Global Change*, Academic Press, San Diego, Calif., 1991.

28. H. W. Jannasch and M. J. Mottl, *Science* **229**, 717–725 (1985).

29. M. Whitfield and D. R. Turner, *Nature* **278**, 132 (1979).

30. M. Whitfield and D. R. Turner, in W. Stumm, ed., *Aquatic Surface Chemistry*, Wiley-Interscience, New York, 1987.

31. D. R. Turner, M. Whitfield, and A. G. Dickson, *Geochim. Cosmochim. Acta* **45**, 855–882 (1981).

32. H. L. Windom, T. Byrd, R. G. Smith, and F. Huan, *Env. Sci. Technol.* **25**, 1137–1142 (1991).

33. A. M. Shiller and E. Boyle, *Nature* **317**, 49–52 (1985).

34. L. Sigg, M. Sturm, W. Stumm, L. Mart, and H. W. Nymberg, *Naturwiss.* **69**, 546–547 (1982).

35. M. M. Shafer and D. E. Armstrong, *Trace Elements in Lake Michigan*, Abstract ACS Meeting, 1990, pp. 273–277.

36. L. Sigg, personal communication, 1993.

37. K. W. Bruland and R. P. Franks, in C. S. Wong, E. Boyle, K. W. Bruland, J. D. Burton, and E. D. Goldberg, eds., *Trace Metals in Sea Water*, Plenum Press, New York, 1983.

38. W. G. Sunda, *Biol. Oceanogr.* **6**, 411–442 (1988–1989).

39. F. J. Millero, *Geochim. Cosmochim. Acta* **59**(4), 661–677 (1995).

40. R. P. Mason, W. F. Fitzgerald, and F. M. M. Morel, *Geochim. Cosmochim. Acta* **58**, 3191–3198 (1994).

41. R. P. Schwarzenbach, D. Imboden, and Ph. M. Gschwend, *Environmental Organic Chemistry*, Wiley-Interscience, New York, 1993.

42. D. Mackay, *Multimedia Environmental Models*, Lewis, Chelsea, MI, 1991.

43. D. A. Dzombak and F. M. M. Morel, *Surface Complexation Modeling: Hydrrous Ferric Oxide*, Wiley-Interscience, New York, 1990.

James J. Morgan
California Institute of Technology

Werner Stumm
Swiss Federal Institute of Technology

POLLUTION

Water is omnipresent on the earth. Constant circulation of water from the ocean to the atmosphere (evaporation) and from the atmosphere to land and the oceans (precipitation, runoff, etc) is generally known as the hydrologic cycle (see Fig. 1) (1–2). Within the hydrologic cyclic, there are several minor and local subcycles where water is used and returned to the environment.

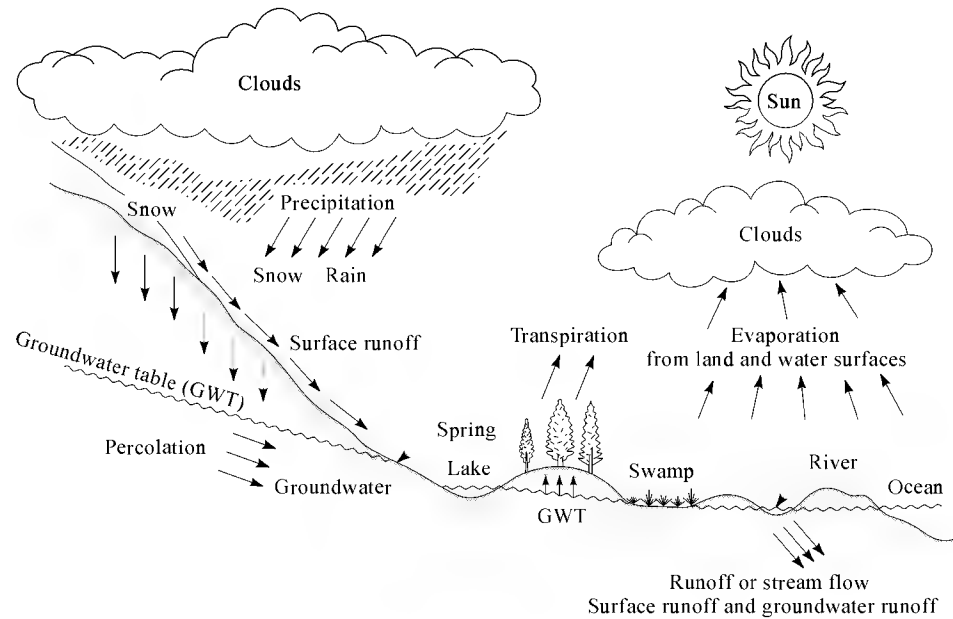


Fig. 1. Schematic diagram of the hydrologic cycle.

The volume of the freshwater amounts to only one-thirtieth of the $1.25 \times 10^9 \text{ km}^3$ ($300 \times 10^6 \text{ mi}^3$) of the water in salty oceans. Approximately one-third of the freshwater exists permanently as snow and ice (3). A large portion of the remaining freshwater has infiltrated too far underground or is partially polluted with minerals and chemicals and therefore is not readily usable. The entire life system on the earth depends on the remaining freshwater sources; therefore, it is essential to protect the quality of the available waters.

The principal hydrologic parameters involved in the storage and transport of freshwater were studied as early as 1894 and are summarized in Table 1 (2).

Table 1. Principal Hydrologic Parameters^a

Storage	Transport
atmospheric water	evaporation from land surface; oceans, lakes, and other water bodies; and snow and ice
oceans and seas	precipitation
rivers	runoff
lakes and reservoirs	infiltration
swamps	
biological water	
soil moisture and groundwater	
frozen water (ice and snow)	

^a Ref. 3.

Freshwater is withdrawn from various sources (rivers, lakes, groundwater, etc) and used many times before its discharge to the ocean. Water uses can generally be classified as follows: public water supply (domestic); industrial; commercial and institutional, eg, restaurants, schools; agricultural; and livestock.

These applications require withdrawal of water from a source and subsequent treatment and conveyance to the point of use. Water is also used without being withdrawn from a source, eg, for navigation, recreation, wild and aquatic life propagation, hydroelectric-power generation, and waste assimilation and transport. The principal types of withdrawal uses and their average rates are given in Table 2. Some of these withdrawal rates represent multiple uses of the same water along main rivers in metropolitan and industrialized areas.

Table 2. Average Daily Water Requirements for Categories of Withdrawal Use in the United States^a

Use	Average daily requirement ^b , $10^3 \text{ m}^3 \text{ d}$ (mgd) ^c
public ^d	102.2 (27,000)
rural	
domestic	9.84 (2,600)
livestock	7.19 (1,900)
agriculture	492.10 (130,000)

industrial	794.90 (210,000)
^a Ref. 4.	
^b Estimated 206 × 10 ⁶ people served.	
^c mgd = million (10 ³) gallons per day.	
^d Residential and municipal (domestic).	

Water-use data and withdrawal rates for public water-supply systems are well documented by municipalities. The U.S. Public Health Service (USPHS) (5), American Water Works Association (AWWA) (6–10), and Federal Housing Administration (FHA) (11) have compiled statistics at regular intervals for >23,000 municipalities in the United States.

The largest consumers of water in the United States are thermal power plants (eg, steam and nuclear power plants) and the iron and steel, pulp and paper, petroleum refining, and food-processing industries. They consume >60% of the total industrial water requirements (see also POWER GENERATION; WASTES, INDUSTRIAL).

Water Quality Management

Over the past decade, water pollution control has progressed from an art to a science. Increased emphasis has been placed on the removal of secondary pollutants, such as nutrients and refractory organics, and on water reuse for industrial and agricultural purposes. This in turn has generated both fundamental and applied research, which has improved both the design and operation of wastewater treatment facilities.

Solving water pollution problems today involves a multidisciplinary approach in which the required water quality is related to agricultural, municipal, recreational, and industrial requirements. In many cases, a cost–benefit ratio must be established between the benefit derived from a specified water quality and the cost of achieving that quality.

Wastewaters emanate from four primary sources: municipal sewage, industrial wastewaters, agricultural runoff, and stormwater and urban runoff. Estimating municipal wastewater flows and loadings can be done in one of several ways, based on knowledge of past and future growth plans for the community, sociological patterns, and land use planning. Two possible ways are as follows: (1) *Population Prediction Techniques*. Several mathematical techniques are available for estimating population growth. Caution should be employed in the use of these procedures, particularly in areas subject to rapid industrial expansion, rapid suburban development, and changing land use patterns; (2) *Saturation Population from Zoning Practice*. Percentages of a saturation population can be estimated for fully developed areas based on zoning restrictions (single-dwelling residential, multiple-dwelling residential, commercial, etc).

Provisions should be included for infiltration in the case of separate sewers as well as storm flows in the case of combined sewers.

As municipal and industrial wastewaters receive treatment, increasing emphasis is being placed on the pollutorial effects of urban and agricultural runoff. The range of concentration of pertinent characteristics in these wastewaters is given in Table 3. Present research on stormwater treatment considers large holding basins in which the stormwaters are treated in the municipal facility after the storm (an *in situ* treatment by screening, sedimentation, chlorination, etc). In the future, water quality management in highly urbanized areas will have to consider stormwater as a primary pollutant.

Table 3. Pollution from Urban and Agricultural Runoff

Constituent	Urban runoff ^a (stormwater)	Agricultural runoff ^b
suspended solids, mg/L	5–1200	
chemical oxygen demand (COD), mg/L	20–610	
biological oxygen demand (BOD), mg/L	1–173	
total phosphorus, mg/L	0.02–7.3	0.10–0.65
nitrate nitrogen, mg/L		0.03–5.00
total nitrogen, mg/L	0.3–7.5	0.50–6.50
chlorides, mg/L	3–35	

^a From Ref. 12.
^b From Ref. 13.

Agricultural runoff is a large contributor to eutrophication in lakes and other natural bodies of water. Effective control measures have yet to be developed for this problem. Runoff of pesticides is also receiving increasing attention.

Water Quality Standards. Water quality standards are usually based on one of two primary criteria, stream standards or effluent standards. Stream standards are based on dilution requirements for the receiving water quality based on a threshold value of specific pollutants or a beneficial use of the water. Effluent standards are based on the concentration of pollutants that can be discharged or on the degree of treatment required.

Stream standards are usually based on a system of classifying the water quality based on the intended use of the water.

Although stream standards are the most realistic in light of the use of the assimilative capacity of the receiving water, they are difficult to administer and control in an expanding industrial and urban area. The equitable allocation of pollutorial loads for many industrial and municipal complexes also poses political and economic difficulties. A stream standard based on minimum dissolved oxygen at low stream flow intuitively implies a minimum degree of treatment. One variation of stream standards is the specification of a maximum concentration of a pollutant (ie, the BOD) in the stream after mixing at a specified low flow condition.

Note that the maintenance of water quality and hence stream standards are not static, but subject to change with the municipal and industrial environment. For example, as the carbonaceous organic load is removed by treatment, the detrimental effect of nitrification in the receiving water increases. Eutrophication may also become a serious problem in some cases. These considerations require an upgrading of the required degree of treatment.

Effluent standards are based on the maximum concentration of a pollutant (mg/L) or the maximum load (lb/day) discharged to a receiving water. These standards can be related to a stream classification.

In 1972 the U.S. Legislature passed Public Law 92-500, which requires certain levels of treatment for industrial wastewater discharges. Effluent guideline criteria (expressed as kilograms pollutant per unit of production) have been developed for each industrial category to be met by specified time periods.

The BPT is defined as the level of treatment that has been proven to be successful for a specific industrial category and that is currently in full-scale operation. Sufficient data exist for this level of treatment so that it can be designed and operated to achieve a level of treatment consistently and with reliability. For example, in the pulp and paper industry, BPT has been defined as biological treatment using the aerated lagoon or the activated sludge process with appropriate pretreatment.

The BAT is defined as the level of treatment beyond BPCTCA that has been proven feasible in laboratory and pilot studies and that is, in some cases, in full-scale operation. BAT in the pulp and paper industry may include such processes as filtration, coagulation for color removal, and improved in-plant control to reduce the wasteload constituents.

In general, effluent guidelines are developed by considering an exemplary plant in a specific industrial category and multiplying the wastewater flow per unit production by the effluent quality attainable from the specified BPT process to obtain the effluent limitation in pounds or kilograms per unit of production. The effluent limitations consider both a maximum 30-day average and a 1-day maximum level. In general, the daily maximum is two to three times the 30-day average. For example, the average wastewater flow from an exemplary plant is 30,000 gal ton of production and the average effluent BOD is 30 mg L.

The effluent limitation can then be computed:

$$(30,000 \text{ gal ton}) \times (8.34 \times 10^{-6}) \times (30 \text{ mg L}) = 7.5 \text{ lb ton}$$

It is recognized that the wastewater volume and characteristics from a specific industrial category will depend on such factors as plant age, size, raw materials used and in-plant processing sequences.

The U.S. Environmental Protection Agency (EPA) has also developed pretreatment guidelines for those industrial plants which discharge into municipal sewer systems. In general, compatible pollutants such as BOD, suspended solids, and coliform organisms can be discharged providing the municipal plant has the capability of treating these wastewaters to a satisfactory level. Noncompatible pollutants, such as grease and oil, heavy metals, etc, must be pretreated to specified levels. Rigid limitations have been developed for the discharge of toxic substances to the nation's waterways.

In several cases, such as shellfish areas and aquatic reserves, the usual water quality parameters do not apply because they are nonspecific as to detrimental effects on aquatic life. For example, COD is an overall measure of organic content, but it does not differentiate between toxic and nontoxic organics. In these cases, a species diversity index has been employed as related to either free-floating or benthic organisms. The index indicates the overall condition to the aquatic environment. It is related to the number of species in the sample. The higher the species diversity index, the more productive the aquatic system. The species diversity index K_D is computed by the equation $K_D = (S - 1) \log_{10} l$, where S is the number of species and l the total number of individual organisms counted.

Regulations establishing effluent limitations guidelines, pretreatment standards and new source performance standards for the organic chemicals, plastics, and synthetic fibers (OCPSF) were promulgated in 1987. In these regulations, specific organic chemicals are defined by the EPA as priority pollutants (see Table 4).

Table 4. Organic Compounds Specified as Pollutants by the EPA

acenaphthene	4-nitrophenol
acrolein	2,4-dinitrophenol
acrylonitrile	4,6-dinitro- <i>o</i> -cresol
benzene	nitrosamines
benzidine	<i>N</i> -nitrosodimethylamine
carbon tetrachloride (tetrachloromethane)	<i>N</i> -nitrosodiphenylamine
	<i>N</i> -nitroso-di- <i>n</i> -propylamine
chlorinated benzenes (other than dichlorobenzene)	pentachlorophenol
chlorobenzene	phenol
1,2,4-trichlorobenzene	phthalate esters
hexachlorobenzene	bis(<i>e</i> -ethylhexyl) phthalate
chlorinated ethanes ^a	butyl benzyl phthalate
1,2-dichloroethane	di- <i>n</i> -butyl phthalate
1,1,1-trichloroethane	di- <i>n</i> -octyl phthalate
hexachloroethane	diethyl phthalate
1,1-dichloroethane	dimethyl phthalate
1,1,2-trichloroethane	polynuclear aromatic hydrocarbons (PAHs)
1,1,2,2-tetrachloroethane	benzo(<i>a</i>)anthracene (1,2-benzanthracene)
chloroethane (ethyl chloride)	benzo(<i>a</i>)pyrene (3,4-benzopyrene)
chloroalkyl ethers ^b	3,4-benzofluoranthene
bis(chloromethyl)ether	
bis(2-chloroethyl)ether	benzo(<i>k</i>)fluoranthene (11,12-benzo- fluoranthene)
2-chloroethyl vinyl ether (mixed)	chrysene
chlorinated naphthalene	acenaphthylene
2-chloronaphthalene	anthracene
chlorinated phenols ^{c,d}	benzo(<i>ghi</i>)perylene (1,12-benzoperylene)
2,4,6-trichlorophenol	fluorene
<i>p</i> -chloro- <i>m</i> -cresol	phenanthrene
chloroform (trichloromethane)	
2-chlorophenol	dibenzo(<i>a,b</i>)anthracene (1,2,5,6- dibenzanthracene)
dichlorobenzenes	
1,2-dichlorobenzene	indeno(1,2,3-cd)pyrene(2,3- <i>o</i> - phenylenepyrene)
1,3-dichlorobenzene	pyrene
1,4-dichlorobenzene	tetrachloroethylene
dichlorobenzidine	toluene
3,3'-dichlorobenzidine	trichloroethylene
dichloroethylenes ^e	vinyl chloride (chloroethylene)
1,1-dichloroethylene	pesticides and metabolites
1,2- <i>trans</i> -dichloroethylene	aldrin
2,4-dichlorophenol	dieldrin
dichloropropane and dichloropropene	
1,2-dichloropropane	chlordan (technical mixture and metabolites)
	DDT and metabolites
1,2-dichloropropylene (1,2-dichloro- propene)	4,4'-DDT
2,4-dimethylphenol	4,4'-DDE (<i>p,p'</i> -DDX)
dinitrotoluene	4,4'-DDD (<i>p,p'</i> -TDE)
2,4-dinitrotoluene	endosulfan and metabolites
2,6-dinitrotoluene	α -endosulfan-alpha
1,2-diphenylhydrazine	β -endosulfan-beta
ethylbenzene	endosulfan sulfate
fluoranthene	endrin and metabolites
haloethers ^c	endrin

4-chlorophenyl phenyl ether	endrin aldehyde
4-bromophenyl phenyl ether	heptachlor and metabolites
bis(2-chloroisopropyl) ether	heptachlor
bis(2-chloroethoxy) methane	heptachlor epoxide
halomethanes ^c	hexachlorocyclohexane (all isomers)
methylene chloride (dichloromethane)	α-BHC-alpha
methyl chloride (chloromethane)	β-BHC-beta
methyl bromide (bromomethane)	γ-BHC (lindane)-gamma
bromoform (tribromomethane)	δ-BHC-delta
dichlorobromomethane	polychlorinated biphenyls (PCBs)
trichlor ofluoromethane	PCB-1242 (Arochlor 1242)
dichlorodifluoromethane	PCB-1254 (Arochlor 1254)
chlorodibromomethane	PCB-1221 (Arochlor 1221)
hexachlorobutadiene	PCB-1232 (Arochlor 1232)
hexachlorocyclopentadiene	PCB-1248 (Arochlor 1248)
isophorone	PCB-1260 (Arochlor 1260)
naphthalene	PCB-1016 (Arochlor 1016)
nitrobenzene	toxaphene
nitrophenols ^f	
2-nitrophenol	2,3,7,8-tetrachlorodibenzo- <i>p</i> -dioxin (TCDD)

^a Including 1,2-dichloroethane, 1,1,1-trichloroethane, and hexachloroethane.
^b Chloromethyl, chloroethyl, and mixed ethers.
^c Other than those listed elsewhere.
^d Including trichlorophenols and chlorinated aresols.
^e 1,1-Dichloroethylene and 1,2-dichloroethylene.
^f Including 2,4-dinitrophenol and dinitrocresol.

These chemicals are regulated as a concentration level in the effluent. In most cases, these levels are in the microgram-per-liter range. Recent air pollution regulations limit the amount of volatile organic carbon (VOC) that can be discharged from wastewater treatment plants. Benzene is a particular case in which air emission controls are required if the concentration of benzene in the influent wastewater exceeds 10 mg L.

Pollution Control

In past years, municipal wastewaters were treated to improve their appearance and bacteriological safety. Treatment included reduction in biochemical oxygen demand (BOD), suspended solids, pathogens, and inorganic dissolved solids. Before 1960, the activated-sludge process, developed in the U.K. in 1914, was used (17). It was based on the capabilities of microorganisms to assimilate organic compounds through oxidative and respiratory mechanisms. The organisms involved could be flocculated and settled in conventional sedimentation vessels (clarifiers). The effluent was relatively clear and required only disinfection before discharge to receiving waters. Preliminary treatment removed grit, sand, and floating debris. It was followed by primary clarification for the removal of suspended solids, grease, and scum. The effluent from the primary clarifiers, containing colloidal and dissolved organic matter, was subjected to activated-sludge treatment and disinfection (eg, chlorination) before discharge. A typical process flow sheet is given in Figure 2 (18); the process components are shown in Figure 3.

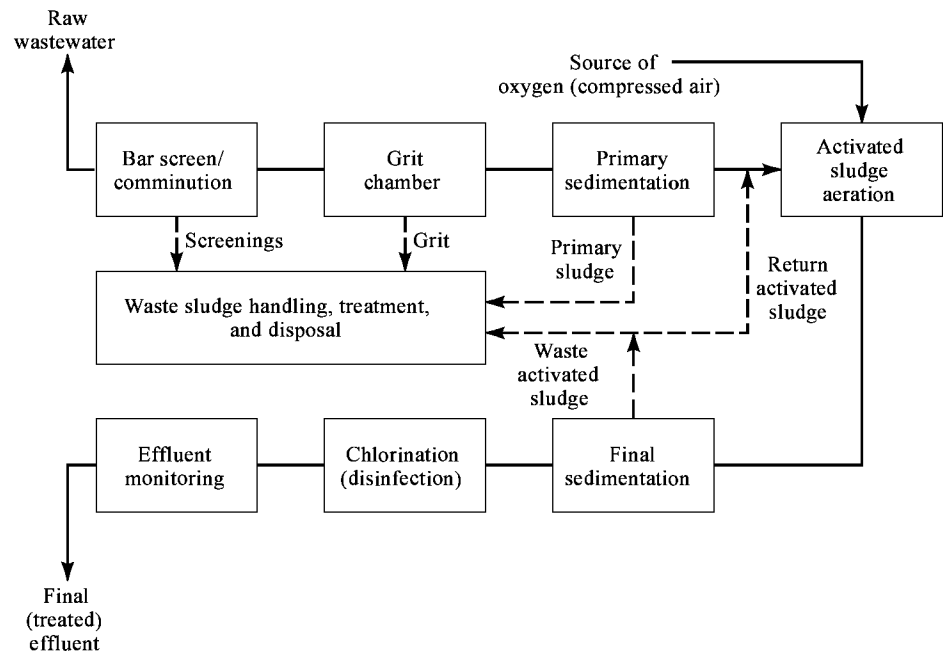


Fig. 2. Typical flow sheet for a domestic wastewater treatment plant utilizing the activated-sludge process.

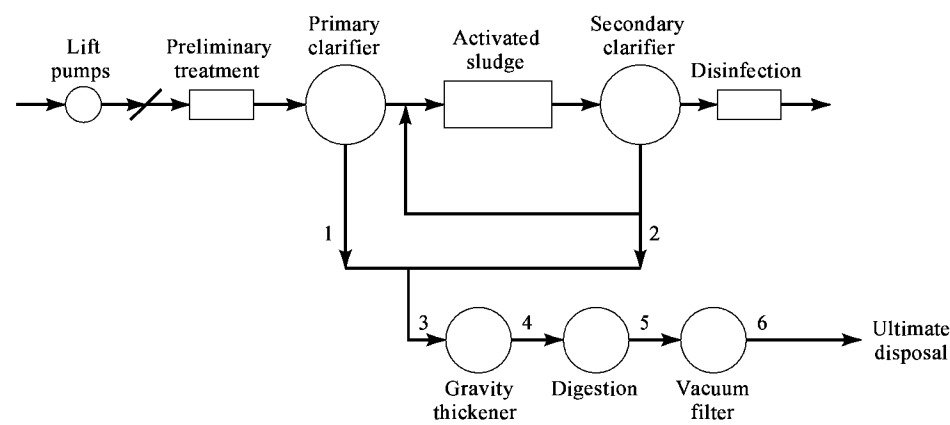


Fig. 3. Process components for the activated-sludge process, with aerobic digestion at plants <473.1 m³ h and two-stage anaerobic digestion at plants >473.1 m³ h (39,40). To convert m³ h to mgd, divide by 157.7.

When industrial wastewaters are mixed with municipal wastes, as in many urban systems, toxic and inhibitory materials are removed in the pretreatment system where nutrient chemicals, eg, nitrogen and phosphorus, are added.

Biological processes (eg, activated sludge, trickling filters, etc) generate primary and biological sludges. These by-products require further treatment and processing before they are suitable for ultimate disposal. In general, the waste sludge is thickened, stabilized by anaerobic or aerobic digestion, and dewatered before land application or land disposal (19,20). The design parameters for the selection and sizing of various process units are given in Table 5 and Figure 3.

Table 5. Design Parameters for the Activated-Sludge Process^a

Parameter, mg L	Influent ^b	Effluent
BOD ₅ ^c	210	20
COD ^d	400	45
TSS ^e	230	20
total-P	11	7
NH ₃ -N	0	0
UOD ^f	406	107

^a Refs. 19,20.

^b Aerobic digestion at plants <473.1 m³ h (3 mgd) and two-stage anaerobic digestion at plants >473.1 m³ h.

^c 5-d Biochemical oxygen demand.

^d Chemical oxygen demand.

^e Total suspended solids.

^f Ultimate oxygen demand.

Before 1960, research and development work was directed toward the improvement of aerobic biological treatment systems. Fixed media were utilized on which microorganisms could attach and grow, eg, stone or plastic trickling filters, rotating biological contactors, etc (see Fig. 4) (21,22). Other designs of attached-growth biological systems include the use of synthetic and redwood media, as well as granular media (Figs. 5 and 6).

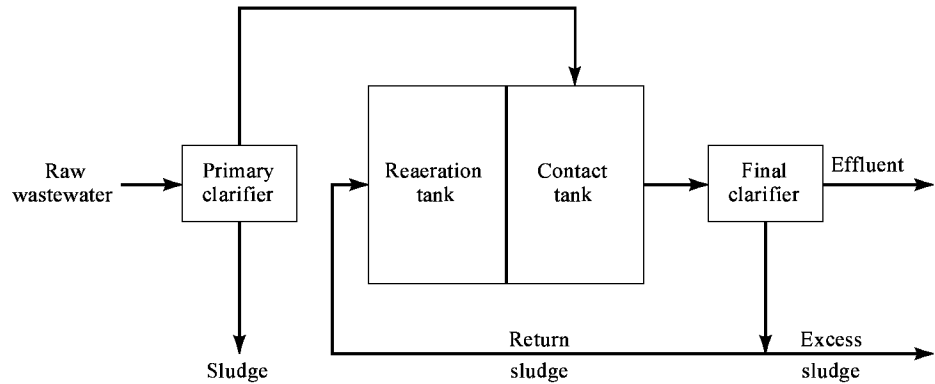


Fig. 4. Contact stabilization plant. Conventional activated-sludge process. The reaeration and contact tanks can be replaced by an aeration tank (21).

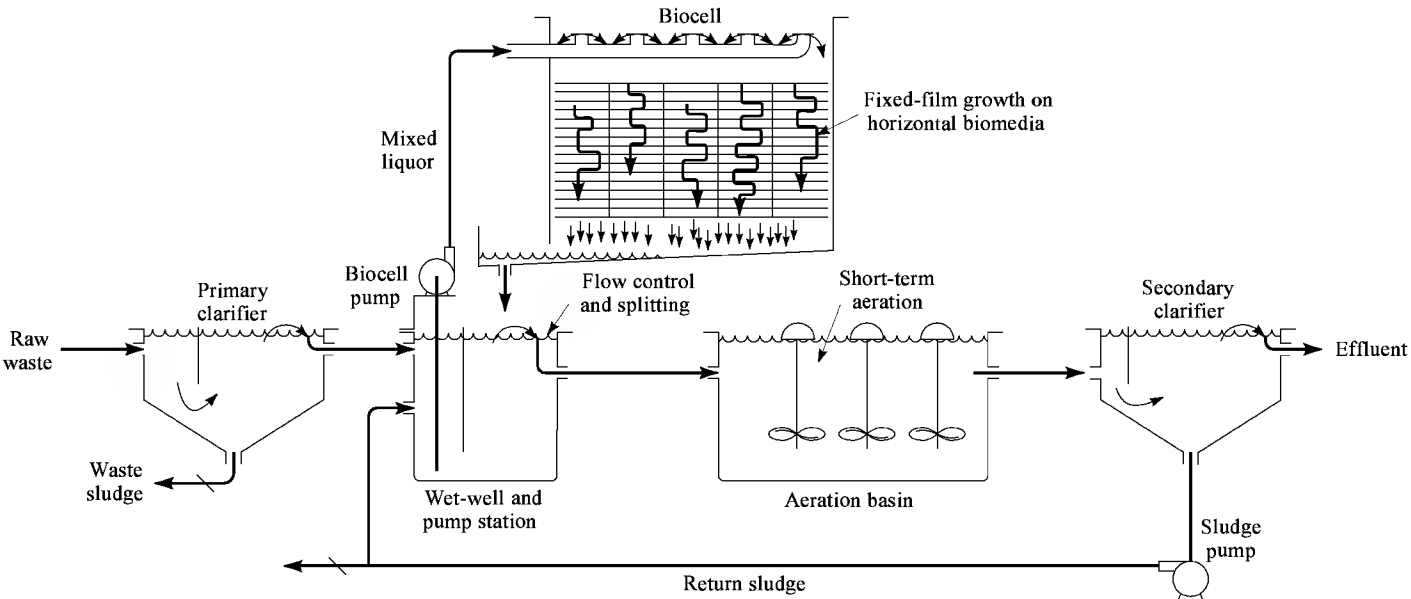


Fig. 5. ABF (activated bio filter) process flow diagram (redwood medium) (22,23).

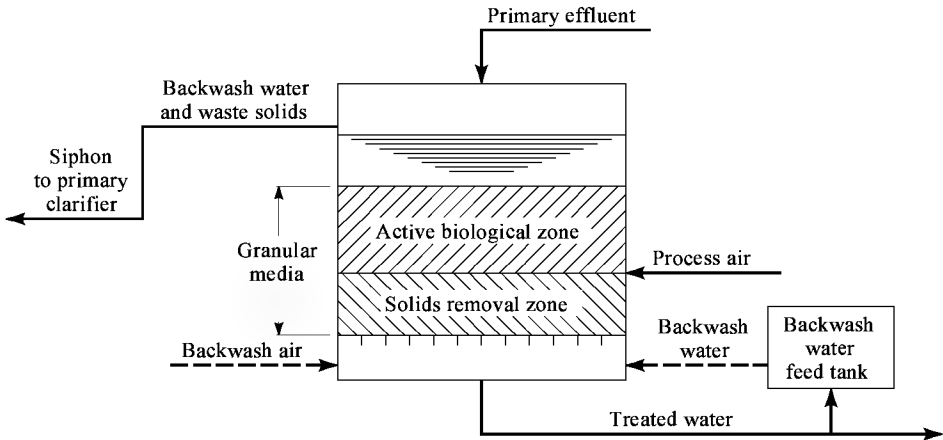


Fig. 6. Biological aerated filter.

Courtesy of Envirotech, Inc.

The post–World War II growth in industrial activity has significantly altered the composition of wastewaters in urban treatment facilities. Pollutants that are resistant to biological oxidation have become predominant (eg, synthetic detergents, petrochemicals, synthetic rubber, etc), requiring the development of new nonbiological processes and approaches to water-pollution control. Today, the industrial-wastewater engineer must be familiar with the manufacturing process and the chemistry of the raw materials, products, and by-products.

Primary Investigation. The basic systems engineering approach is the most suitable method for developing a solution to industrial wastewater management. The preliminary investigation includes a plant survey and the characterization of the wastewater source.

A wastewater survey provides the facts and data necessary to complete a wastewater management plan which comprises the following steps: segregation of clean water for potential recovery and reuse; isolation and segregation of noncompatible waste streams for separate pretreatment (eg, ion-exchange-regeneration wastes, inorganic and organic waste streams, waste streams that contain potentially toxic compounds, etc); isolation and segregation of concentrated waste streams and streams containing solvents for possible recovery or separate treatment and disposal (ie, thermal decomposition with heat recovery); and identification and characterization of batch discharges and spills in order to incorporate protective systems (eg, equalization) for the treatment facility.

Point number	Sludge, kg m ³	Concentration, %
1	129.4	4
2	98.3	0.8
3	227.7	2.6
4	227.7	8
5	113.8	5
6	113.8	20

In-Plant Waste Control. Pollution can be reduced or eliminated by process modification, chemical and raw materials substitution, or recovery of by-products. In addition, process modification generally increases product yield by incorporating control devices.

Waste Minimization. Before end-of-pipe wastewater treatment or modifications to existing wastewater treatment facilities to meet new effluent criteria, a program of waste minimization should be initiated.

Reduction and recycling of waste are inevitably site and plant specific, but a number of generic approaches and techniques have been used successfully across the country to reduce many kinds of industrial wastes.

Generally, waste minimization techniques can be grouped into four major categories: inventory management and improved operations, modification of equipment, production process changes, and recycling and reuse. Such techniques can have applications across a range of industries and manufacturing processes and can apply to hazardous as well as nonhazardous waste.

Many of these techniques involve source reduction—the preferred option on the EPA's hierarchy of waste management (24). Others deal with on- and off-site recycling. The best way to determine how these general approaches can fit a particular company's needs is to conduct a waste minimization

assessment, as discussed above. In practice, waste minimization opportunities are limited only by the ingenuity of the generator. In the end, a company looking carefully at bottom-line returns may conclude that the most feasible strategy would be a combination of source reduction and recycling projects.

Waste minimization approaches as developed by the U.S. EPA are as follows:

Inventory Management and Improved Operations. (1) Inventory and trace all raw materials. (2) Purchase fewer toxic and more nontoxic production materials. (3) Implement employee training and management feedback. (4) Improve material receiving, storage, and handling practices.

Modification of Equipment. (1) Install equipment that produces minimal or no waste. (2) Modify equipment to enhance recovery or recycling options. (3) Redesign equipment or production lines to produce less waste. (4) Improve operating efficiency of equipment. (5) Maintain strict preventive maintenance program.

Production Process Changes. (1) Substitute nonhazardous for hazardous raw materials. (2) Segregate wastes by type for recovery. (3) Eliminate sources of leaks and spills. (4) Separate hazardous from nonhazardous wastes. (5) Redesign or reformulate end products to be less hazardous. (6) Optimize reactions and raw material use.

Recycling and Reuse. (1) Install closed-loop systems. (2) Recycle on site for reuse. (3) Recycle off site for reuse. (4) Exchange wastes.

In order to implement the program, an audit needs to be made, as described in the following:

Phase I: Precassessment. (1) Audit focus and preparation. (2) Identify unit operations and processes. (3) Prepare process flow diagrams.

Phase II: Mass Balance. (1) Determine raw material inputs. (2) Record water usage. (3) Assess present practice and procedures. (4) Quantify process outputs. (5) Account for emissions: to atmosphere, to wastewater, and to off-site disposal. (6) Assemble input and output information. (7) Derive a preliminary mass balance. (8) Evaluate and refine the mass balance.

Phase III: Synthesis. (1) Identify options: identify opportunities, target problem areas, and confirm options. (2) Evaluate options: technical, environmental, and economic. (3) Prepare action plan: waste reduction plan, production efficiency plan, and training.

For further information see WASTES, INDUSTRIAL and Ref. 16.

Toxic Organic Materials. The term toxic organics includes synthetic organic compounds such as pesticides, herbicides, PCBs, and chlorinated hydrocarbons, usually produced by the manufacturers and formulators of these products.

Because these compounds persist over a long period of time in a natural environment, the most effective treatment technology at present is incineration. The recommended temperatures for incineration of chlorinated hydrocarbons and pesticides range from 982 to 1482°C. A sustained high temperature prevents the emission of degradation products. The incinerator stack gases generally contain HCl vapors that require the installation of scrubbers. In general, a vortex burner provides satisfactory performance. The mixture is preheated, the liquid vaporized, and the gases heated to ignition temperature (28,29) (see INCINERATORS).

The vortex burner maintains stable combustion temperature when the organic concentration in the waste is sufficiently high and has a heating value of ca 10.5–12.6 MJ kg (4500–5400 Btu lb). Auxiliary fuel may be required when the chloride concentration in the waste exceeds 70% (30).

In wet-air oxidation, the aqueous mixture is heated under pressure in the presence of air, which oxidizes the organic material. The efficiency of the oxidation process is a function of reaction time and temperature. The oxidation products are generally less complex and can be treated by conventional biological methods (31). The reactor usually operates between 177 and 321°C with pressures of 2.52–20.8 MPa (350–3000 psig).

If the concentration of organic material is too low, the following technique discussed below may be used.

Membrane Separation. Reverse osmosis (qv) or ultrafiltration (qv) can be used to concentrate toxic organic substances, depending on the type of compounds and the stability of the membrane against chemical attack (see also MEMBRANE TECHNOLOGY). In general, high molecular weight compounds have higher rejection rates than those obtained with lower molecular weight compounds. Typical rejection rates are given in Table 6.

Table 6. Membrane Rejection Rates

Compound	Percent
lindane	84
DDT ^a	99.5
DDD ^b	99.9
2,4-D ^c	99.9
chlorinated hydrocarbons	98.95–100
organophosphorus pesticides	98.05–98.88
miscellaneous pesticides	72–100

^a Dichlorodiphenyltrichloroethane.

^b 1,1-Dichloro-2,2-bis(*p*-chlorophenyl)ethane.

^c 2,4-Dichlorophenoxy acetic acid.

Chemical Treatment. Some organic compounds are attacked by chemical reagents such as potassium permanganate, sodium hydroxide, calcium hypochlorite, and ozone (29,30).

Potassium permanganate oxidizes heptachlor with ca 80–90% efficiency (30). Sodium hydroxide degrades malathion, lindane, and DDT (30,32). Ozone oxidizes dissolved organic compounds, including toxic substances, because of an oxidation potential higher than that of permanganate, hydrogen peroxide, or hypochlorite. The ozonation system must be designed to carry the reaction to completion in order to prevent the generation of toxic intermediates. Ultraviolet radiation in conjunction with ozone is a highly effective degradation technique for malathion, DDT, pentachlorophenol, dichlorobutane, dichlorobenzene, PCBs, and chloroform (33,34).

Ozone alone oxidizes phenolic compounds to carbon dioxide and water. However, at concentrations of 1.5–2.5 parts ozone per part of phenol, the phenolic compounds can be converted to less toxic and biodegradable intermediates in a cost-effective manner (35).

Adsorption. Organic compounds are adsorbed on activated carbon and synthetic resins (eg, XAD-2 and XAD-4, Rohm and Haas Co.). This technique depends on the properties of the compound being removed and the regenerative capability of the adsorbent. The EPA has developed carbon-adsorption isotherms for various toxic organic compounds, and the results are shown in Table 7 (36). The following compounds are not adsorbed on activated carbon: acetone cyanohydrin, butylamine, choline chloride, cyclohexylamine, diethylene glycol, ethylenediamine, hexamethylenediamine, morpholine, and triethanolamine.

Table 7. Adsorption Capacities of Activated Carbon

Compound	Adsorption capacity, mg g	Compound	Adsorption capacity, mg g
bis(2-ethylhexyl) phthalate	11,300	guanine	120
butyl benzyl phthalate	1,520	styrene	120
heptachlor	1,220	1,3-dichlorobenzene	118
heptachlor epoxide	1,038	acenaphthylene	115
endosulfan sulfate	686	4-chlorophenyl phenyl ether	111
endrin	666	diethyl phthalate	110

fluoranthene	664	2-nitrophenol	99
aldrin	651	dimethyl phthalate	97
PCB- 1232	630	hexachloroethane	97
β-endosulfan	615	chlorobenzene	91
dieldrin	606	<i>p</i> -xylene	85
hexachlorobenzene	450	2,4- dimethylphenol	78
anthracene	376	4-nitrophenol	76
4-nitrobiphenyl	370	acetophenone	74
fluorene	330		74
DDT	322	1,2,3,4-tetrahydronaph-thalene	
2-acetylaminofluorene	318	adenine	71
α-BHC	303	dibenzo[<i>a,b</i>]anthracene	69
anethole	300	nitrobenzene	68
3,3-dichlorobenzidine	300	3,4-benzofluoranthene	57
2-chloronaphthalene	280	1,2-dibromo-3-chloropropane	53
phenylmercuric acetate	270	ethylbenzene	53
hexachlorobutadiene	258	2-chlorophenol	51
γ-BHC (lindane)	256	tetrachloroethylene	51
<i>p</i> -nonylphenol	250	<i>o</i> -anisidine	50
4-dimethylaminoazobenzene	249	5-bromouracil	44
chlordane	245	benzo[<i>a</i>]pyrene	34
PCB- 1221	242	2,4- dinitrophenol	33
	232	isophorone	32
1,1-dichloro-2,2-bis(<i>p</i> -chloro- phenyl)ethane (DDE)		trichloroethylene	28
acridine yellow	230	thymine	27
benzidine dihydrochloride	220	toluene	26
β-benzenehexachloride	220	5-chlorouracil	25
<i>n</i> -butyl phthalate	220	<i>N</i> -nitrosodi- <i>n</i> -propylamine	24
<i>N</i> -nitrosodiphenylamine	220	bis(2-chloroisopropyl) ether	24
phenanthrene	215	phenol	21
dimethylphenylcarbinol	210	bromoform	20
4-aminobiphenyl	200	carbon tetrachloride	11
2-naphthol	200	bis(2-chloroethoxy)methane	11
α-endosulfan	194	uracil	11
acenaphthene	190	benzo[<i>ghi</i>]perylene	11
	190	1,1,2,2-tetrachloroethane	11
4,4'-methylenebis(2-chloro-aniline)		1,2-dichloropropene	8.2
benzo[<i>k</i>]fluoranthene	181	dichlorobromomethane	7.9
acridine orange	180	cyclohexanone	6.2
1-naphthol	180	1,2-dichloropropane	5.9
4,6-dinitro- <i>o</i> -cresol	169	1,1,2-trichloroethane	5.8
1-naphthylamine	160	trichlorofluoromethane	5.6
2,4-dichlorophenol	157	5-fluorouracil	5.5
1,2,4-trichlorobenzene	157	1,1-dichloroethylene	4.9
2,4,6-trichlorophenol	155	dibromochloromethane	4.8
2-naphthylamine	150	2-chloroethyl vinyl ether	3.9
pentachlorophenol	150	1,2-dichloroethane	3.6
2,4-dinitrotoluene	146	1,2- <i>trans</i> -dichloroethylene	3.1
2,6-dinitrotoluene	145	chloroform	2.6
4-bromophenyl phenyl ether	144	1,1,1-trichloroethane	2.5
<i>p</i> -nitroaniline	140	1,1-dichloroethane	1.8
1,1-diphenylhydrazine	135	acrylonitrile	1.4
naphthalene	132	methylene chloride	1.3
1-chloro-2-nitrobenzene	130	acrolein	1.2
1,2-dichlorobenzene	129	cytosine	1.1
<i>p</i> -chloro- <i>m</i> -cresol	124	benzene	1.0
1,4-dichlorobenzene	121	EDTA	0.86
benzothiazole	120	benzoic acid	0.76
diphenylamine	120	chloroethane	0.59
		<i>N,N</i> -dimethylnitrosamine	6.8 × 10 ⁻⁵

Synthetic polymeric adsorbents have a high porosity, large surface area, and an inert hydrophobic surface. These resins can be regenerated chemically, which produces a concentrated waste stream requiring further treatment or disposal. The adsorption capacity of the polymeric adsorbent XAD-4 for a group of chlorinated hydrocarbons is given in Table 8. The EPA may recommend a combination of air stripping and carbon adsorption wherein air stripping removes most of the volatile organics and adsorption removes the rest (37).

Table 8. Adsorption Capacity of Amberlite XAD-4 Polymeric Resin ^a

Compounds	Maximum solubility in water, g /L	Influent concentration, mg /L	Adsorption capacity, kg /m ³
phenol	82	250	12.5
2-chlorophenol	26	350	38.5
2,4-dichlorophenol	4.5	430	81.6
2,4,6-trichlorophenol	0.9	510	192.3

^a Ref. 36.

Heavy Metals. Heavy metals of particular concern in the treatment of wastewaters include copper, chromium, zinc, cadmium, mercury, lead, and nickel. They are usually present in the form of organic complexes, especially in wastewaters generated from textiles finishing and dye chemicals manufacture.

Inorganic heavy metals are usually removed from aqueous waste streams by chemical precipitation in various forms (carbonates, hydroxides, sulfide) at different pH values. The solubility curves for various metal hydroxides, when they are present alone, are shown in Figure 7. The presence of other metals and complexing agents (ammonia, citric acid, EDTA, etc) strongly affects these solubility curves and requires careful evaluation to determine the residual concentration values after treatment (see Table 9) (38,39).

Table 9. Typical Residual Concentrations after Chemical Precipitation^a

Metal	Agent	Residual concentration, mg/L
cadmium	soda ash	0.3
chromium (hexavalent)	sodium bisulfite and lime	0.05
chromium (total)	caustic, lime	0.5
copper	caustic, lime	0.5
nickel	soda ash	0.5
zinc	caustic, lime	0.5

^a Ref. 38.

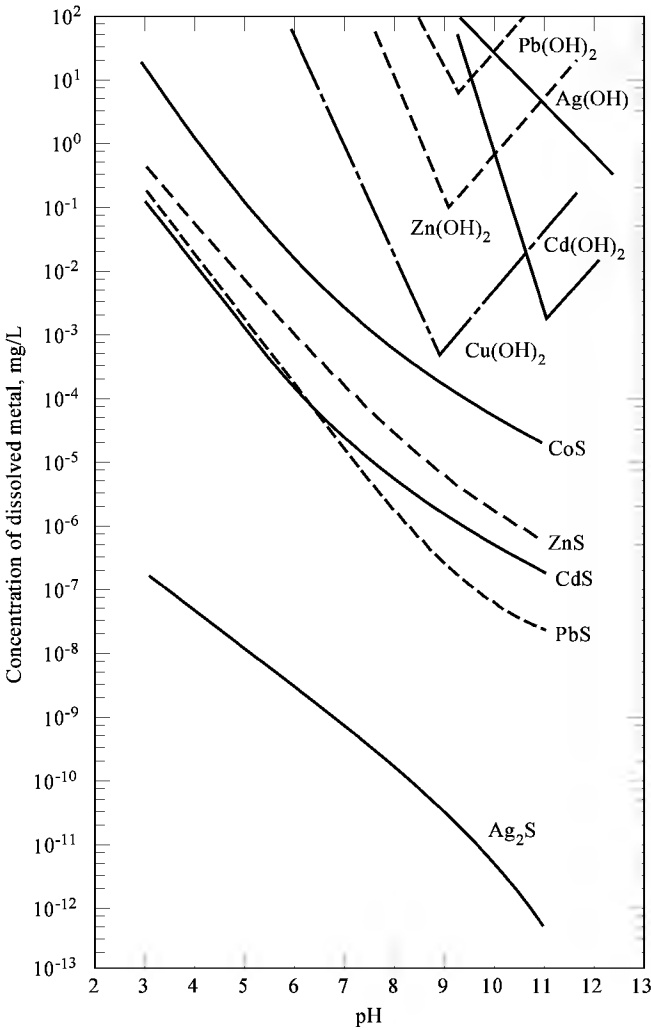


Fig. 7. Theoretical solubilities of metal hydroxides and sulfides as a function of pH (37).

Other methods, including activated carbon, ion exchange, and reverse osmosis, can be used to concentrate waste streams and remove the heavy metals. Activated carbon is effective in reducing hexavalent chromium, mercury, and many metals complexed by organic liquids. Similarly, various ion-exchange resins have been found to be effective in reducing metal ions from solution. The spent resins and activated carbon may require chemical regeneration, however, which may produce a concentrated waste stream that again requires treatment.

See WASTES, INDUSTRIAL for more information on wastewater treatment.

BIBLIOGRAPHY

"Water, Pollution," in *ECT* 3rd ed., Vol. 24, pp. 295–314, by Muthiah Ramanathan, Roy F. Weston, Inc.

1. R. K. Linsley and J. B. Franzini, *Water Resources Engineering*, McGraw-Hill Book Company, New York, 1979.
2. G. M. Fair, J. C. Geyer, and D. A. Okun, *Water and Wastewater Engineering*, John Wiley & Sons, Inc., New York, 1966.
3. D. K. Todd, *Groundwater Hydrology*, John Wiley & Sons, Inc., New York, 1980.
4. H. W. Gehm and J. I. Bregman, *Handbook of Water Resources and Pollution Control*, Van Nostrand Reinhold Company, New York, 1976.
5. J. H. Feth, *Water Facts and Figures for Planners and Managers*, Publication No. 24, U.S. Public Health Service, Washington, D.C., 1962.
6. American Water Works Association, *A Survey of Operating Data for Water Works in 1960*, AWWA staff report, New York, 1964.
7. American Water Works Association, *Supplement to a Survey of Operating Data for Water Works in 1960*, AWWA staff report, New York, 1964.
8. B. R. Beattie and H. S. Forster, Jr., *Proceedings of the 1979 Annual AWWA Conference*, American Water Works Association, New York, 1979.
9. D. W. Browne, F. Minton, and C. Barnhill, *J. Am. Water Works Assoc.* **72**(9), 506 (1980).
10. American Water Works Association, *AWWA Committee Report on Water Use*, New York, 1972.
11. F. P. Lineweaver, Jr., J. C. Geyer, and J. B. Wolff, *A Study of Residential Water Use*, Federal Housing Administration, U.S. Department of Housing

and Urban Development, Washington, D.C., 1967.

12. Metcalf and Eddy, *Wastewater Engineering Treatment, Disposal and Reuse*, 3rd ed., McGraw-Hill, New York, 1991.

13. D. L. Ford, "Application of the Total Carbon Analyzer for Industrial Wastewater Evaluation" in *Proceedings 23rd Industrial Waste Conference*, Purdue University, 1968.

14. P. W. Lankford and W. W. Eckenfelder, eds., *Toxicity Reduction in Industrial Effluents*, Van Nostrand Reinhold, New York, 1990.

15. J. W. Patterson, *Industrial Wastewater Treatment Technology*, Butterworth, Boston, 1985.

16. W. W. Eckenfelder, Jr., "Water Quality Management," in A. Bisio and S. Boots, eds., *Encyclopedia of Energy Technology and the Environment*, Vol. 4, John Wiley & Sons, Inc., New York, 1995, pp. 2823–2859.

17. H. W. Van Gills, *Bacteriology of Activated Sludge*, report no. 32-IG-TNO, Research Institute for Public Health Engineering, the Hague, The Netherlands, 1964.

18. Metcalf, Eddy, Inc., *Wastewater Engineering: Collection, Treatment, and Disposal*, McGraw-Hill Book Company, New York, 1972.

19. *Wastewater Treatment Systems, Performance and Cost*, Roy F. Weston, Inc., West Chester, Pa., 1977.

20. U.S. Environmental Protection Agency, *Innovative and Alternative Technology Assessment Manual*, Office of Water Program Operations, Washington, D.C., EPA-430 9-78-009, 1978.

21. Roy F. Weston, Inc., *Process Design Manual for Upgrading Existing Wastewater Treatment Plants*, prepared for U.S. EPA, technology transfer, program no. 17090 GNQ, 1971.

22. M. Ramanathan and W. E. Verley, *Evaluation, Design and Startup of an Innovative and Cost-effective Wastewater Treatment Plant at Concord, New Hampshire*, paper presented at the 36th Annual Meeting, Virginia Water Pollution Control Association, Inc., Charlottesville, Va., 1982.

23. *Activated Biofiltration*, bulletin no. KL-42330-2, Neptune Microfloc, Inc., Corvallis, Ore., 1980.

24. U.S. Environmental Protection Agency (EPA), *USEPA Manual for Waste Minimization Opportunity Assessments*, EPA 600 2-88 025, EPA, Washington, D.C., 1988.

25. U.S. Environmental Protection Agency, *A Handbook of Key Federal Regulations and Criteria for Multimedia Environmental Control*, Industrial Environmental Research Laboratory, Research Triangle Park, N.C., EPA-600 7-79-175, 1979.

26. H. D. Stensel, *Biological Aerated Filter*, presented at the I A Technologies Seminar sponsored by the U.S. Environmental Protection Agency, 1982.

27. U.S. Environmental Protection Agency, *Process Control Manual for Aerobic Biological Wastewater Treatment Facilities*, Office of Water Program Operations, Washington, D.C., EPA-430 9-77-006, 1977.

28. Y. H. Kiang, *Chem. Eng. Progr.* **72**, 71 (1976).

29. J. J. Santoleri, *Chem. Eng. Progr.* **69**, 68 (1973).

30. G. M. Leigh, *J. Water Pollut. Control Fed.* **41**(11), R450 (1969).

31. *Technology and Application—Summary*, Zimpro Environmental Systems, Inc., Rothschild, Wisc., 1978.

32. M. V. Kennedy, B. J. Stojanovic, and F. L. Shuman, Jr., *J. Agric. Food Chem.* **20**(2), (1972).

33. H. W. Prengle, Jr., *Proceedings of the Seminar on Wastewater Treatment and Disinfection with Ozone*, International Ozone Association, Vienna, Va., 1977.

34. R. C. Rice, *AIChE Symp. Ser., Water 1980*, **77**(209), (1981).

35. K. H. Lanouette, *Chem. Eng.* **48**(22), (Oct. 1977).

36. R. A. Dobbs, R. J. Middendorf, and J. M. Cohen, *Carbon Adsorption Isotherms for Toxic Organics*, Municipal Environmental Research Laboratory, U.S. Environmental Protection Agency, Cincinnati, Ohio, EPA-600 8-80-023, 1980.

37. *Chem. Week*, 44 (June 22, 1983).

38. The Permutit Company, *Proceedings of Seminar on Metal Wastes Treatment Featuring the Sulfex Process*, The Permutit Co., Paramus, N.J., 1978.

39. K. H. Lanoutte, *Chem. Eng.* **48**(22), (Oct. 1977).

W. Wesley Eckenfelder, Jr.
Eckenfelder Inc.

ANALYSIS

Since 1970, new analytical techniques, eg, ion chromatography, have been developed, and others, eg, atomic absorption and emission, have been improved (1–5). Detection limits for many chemicals have been dramatically lowered. Many wet chemical methods have been automated and are controlled by microprocessors which allow greater data output in a shorter time. Perhaps the best known continuous-flow analyzer for water analysis is the Autoanalyzer system manufactured by Technicon Instruments Corp. (Tarrytown, N.Y.) (6). Isolation of samples is maintained by pumping air bubbles into the flow line. Recently, flow-injection analysis has also become popular, and a theoretical comparison of it with the segmented flow analyzer has been made (7–9). Several books are available regarding automated chemical analysis (10,11) (see ANALYTICAL METHODS; also BIOTECHNOLOGY).

Although simple analytical tests often provide the needed information regarding a water sample, such as the formation and presence of chloroform and other organohalides in drinking water, require some very specialized methods of analysis. The separation of trace metals into total and uncomplexed species also requires special sample handling and analysis (12).

A list of all water analyses would be extremely long since, under some conditions and with enough time, water can solubilize everything to some extent. Fortunately, a great deal can be learned about a water supply by carrying out a few physical and chemical tests. These simple tests might be all that are needed to characterize a water supply for many purposes, and it is usually the purpose for which the water is to be used that determines the type and extent of testing. The methods described in this review are intended primarily for freshwater analysis and may not be suitable for the analysis of saline water. In addition, there are several books, manuals, and annual review articles available that detail water analysis for different substances and uses (13–21).

Physical Properties

Temperature. Water temperature is an important parameter in calculations of oxygen solubility, calcium carbonate saturation and stability, and various forms of alkalinity, as well as in determining basic hydrobiological characteristics. The temperature should be taken *in situ* for accuracy, and a standard mercury thermometer with readings to the nearest 0.1°C should be used. It should be calibrated against a precision thermometer certified by the National Bureau of Standards. A thermistor is preferable when attempting to measure temperature at different depths and for automated monitoring and surveillance, and should be similarly calibrated (see TEMPERATURE MEASUREMENT).

Specific Conductance. The specific conductance depends on the total concentration of the dissolved ionized substances, ie, the ionic strength of a water sample. It is an expression of the ability of the water to conduct an electric current. Freshly distilled water has a conductance of 0.5–2 μS cm, whereas that of potable water generally is 50–1500 μS cm. The conductivity of a water sample is measured by means of an a-c Wheatstone-bridge circuit with a null indicator and a conductance cell. Each cell has an associated constant which, when multiplied by the conductance, yields the specific conductance.

The concentration of dissolved ionic substances can be roughly estimated by multiplying the specific conductance by an empirical factor of 0.55–0.9, depending on temperature and soluble components. Since specific conductance is temperature dependent, all samples should be measured at the same temperature. Alternatively, an appropriate temperature-correction factor obtained by comparisons with known concentrations of potassium chloride may be used. Instruments are available that automatically correct conductance measurements for different temperatures.

Color. Many water samples have a yellow to brownish-yellow color which is caused by natural substances, eg, leaves, bark, humus, and peat material. Turbidity in a sample can make the measurement of color uncertain and is usually removed by centrifugation prior to analysis. The color is usually measured by comparison of the sample with known concentrations of colored solutions. A platinum–cobalt solution is used as the standard, and the unit of color is that produced by 1 mg L platinum as chloroplatinate ion. The standard is prepared from potassium chloroplatinate (K₂PtCl₆) and cobalt chloride (CoCl₂·6H₂O). The sample may also be compared to suitably calibrated special glass color disks.

Three special tristimulus light filters are available which, when combined with a specific light source and a filter photometer, can be used to obtain color data (20). Although this method provides added precision and accuracy, it is seldom worth the extra effort required (see COLOR).

Turbidity. Turbidity in natural water is caused by the presence of suspended matter which scatters light. The suspended material is usually clay or silt, finely divided organic and inorganic material, or microorganisms. It is measured either visually or by a nephelometric method. The former is based on the light path through a sample that just causes the image of the flame of a standard candle to disappear. Longer light paths are indicative of lower turbidities. Suspensions of clay are used as standards and results are reported in Jackson turbidity units (JTU). The lowest turbidity value that can be measured using this method is 25 JTU.

The nephelometric method involves illuminating the sample in a turbidimeter and measuring the amount of light scattered at 90° to the incident beam. The higher the intensity of scattered light, the greater the sample turbidity. A formazan polymer suspension is used as the standard and results are reported in nephelometric turbidity units (NTU). The greater precision, sensitivity, and wider applicability of this method make it preferable to the visual method.

Taste and Odor. The measurement of taste and odor is somewhat subjective and depends on the personal judgements of individuals. Panels of not less than five observers, and preferably more than ten, are used. The sample is diluted with odor-free water until a ratio at which the odor is just perceptible is determined; this ratio is called the threshold odor number (TON). A similar method is used to detect a distinct taste in water (see FLAVOR CHARACTERIZATION).

Dissolved Solids. Dissolved solids are materials that pass through a glass-fiber filter and remain after evaporation and drying at 180°C.

Suspended Solids. Suspended solids are determined by filtering a known volume of water through a glass-fiber filter and weighing the filter before and after filtration. The filter is dried at 105–110°C, and the weight difference is equal to the suspended solids.

pH. The pH of most natural waters is 4–9 and should be measured *in situ* since it is subject to change once a sample has been isolated. It can be measured either colorimetrically or potentiometrically. The former method relies on color-indicator papers which are impregnated with pH-sensitive dyes. The potentiometric method requires the use of a pH-sensitive glass electrode, which is generally interference-free at pH 4–9. The electrode should be calibrated frequently against standard buffer solutions of known pH (see HYDROGEN-ION ACTIVITY).

Principal Mineral Constituents and Gases

Alkalinity. The alkalinity of a water sample is its acid-neutralizing capacity. Bicarbonate and carbonate ions are the predominant contributors to alkalinity in most waters, and their chemical equilibria generally maintain the pH of 5–9. The presence of enough hydroxide ion to affect the alkalinity determination in natural waters is rare. Silica, borate, or phosphate do contribute to the overall alkalinity if present in large enough quantities.

The alkalinity is determined by titration of the sample with a standard acid (sulfuric or hydrochloric) to a definite pH. If the initial sample pH is >8.3, the titration curve has two inflection points reflecting the conversion of carbonate ion to bicarbonate ion and finally to carbonic acid (H₂CO₃). A sample with an initial pH <8.3 only exhibits one inflection point corresponding to conversion of bicarbonate to carbonic acid. Since most natural-water alkalinity is governed by the carbonate–bicarbonate ion equilibria, the alkalinity titration is often used to estimate their concentrations.

Acidity. Acidity is the base-neutralizing capacity of a sample of water. It is determined by titration of the sample with standard base to pH 8.3

(phenolphthalein end point). Generally, a sample is not reported as acidic unless its initial pH is <4.5. The acidity may be the result of free acids, eg, HCl or H₂SO₄, or the hydrolysis of certain metal cations.

The free-acid content is determined by titration of a cold solution to pH 4.5. The total acidity is determined by titration to pH 8.3 in a boiling solution. Some natural-water samples might be complex, and the best determination of acidity results from visual inspection of the plotted titration curve.

Hardness (Calcium and Magnesium). The term *hardness* has its origins in the household laundry use of water. Some waters were considered harder to use for laundering because they required more soap to produce suds. The hardness of water was then taken to be a measure of the capacity of the water to cause the precipitation of soap. This is actually a result of the reaction of soap with calcium or magnesium ions. Thus, water hardness is generally a measure of the total concentration of calcium and magnesium. The portion of hardness that disappears with boiling is temporary or carbonate hardness and is primarily caused by calcium and magnesium bicarbonates. They precipitate as their carbonates by the loss of carbon dioxide during boiling. The hardness remaining after boiling is the permanent or noncarbonate hardness. Other cations, eg, strontium or barium, also contribute to hardness, but their concentration is usually insignificant (see DISPERSANTS; DETERGENCY).

Calcium and magnesium can be titrated readily with disodium ethylenediaminetetraacetate, with Eriochrome Black T as the indicator. The solution is buffered at pH 10.0. Certain metal ions interfere with this procedure by causing fading or indistinct end points. Cyanide, sulfide, or hydroxylamine can be used to eliminate or minimize the interferences.

Hardness can also be calculated by summation of the individually determined alkaline earths by means of atomic absorption analysis. Basic samples must be acidified, and lanthanum chloride must be added to minimize interferences from phosphate, sulfate, and aluminum. An ion-selective electrode that utilizes a liquid ion exchanger is also available for hardness measurement; however, this electrode is susceptible to interferences from other dissolved metal ions.

Magnesium, calcium, barium, and strontium can also be determined by ion chromatography with *m*-phenylenediamine in perchloric acid as the eluent. Ion chromatography by conductimetric detection has been described, and applications to environmental waters have been discussed (1,22–23).

Sodium and Potassium. Sodium and potassium can be determined by either atomic emission or absorption. Large concentrations of sodium can interfere with the potassium determination in either of these methods. Excess sodium can be added to both the potassium standards and samples to minimize any variations in the samples. Proper positioning of the flame helps reduce sodium interference in atomic absorption.

Chloride. Chloride is common in freshwater because almost all chloride salts are very soluble in water. Its concentration is generally 10^{–4} to 10^{–3} *M*. Chloride can be titrated with mercuric nitrate. Diphenylcarbazone, which forms a purple complex with the excess mercuric ions at pH 2.3–2.8, is used as the indicator. The pH should be controlled to ±0.1 pH unit. Bromide and iodide are the principal interferences, whereas chromate, ferric, and sulfite ions interfere at levels greater than 10 mg/L. Chloride can also be determined by a colorimetric method based on the displacement of thiocyanate ion from mercuric thiocyanate by chloride ion. The liberated SCN[–] reacts with ferric ion to form the colored complex of ferric thiocyanate. The method is suitable for chloride concentrations from 10^{–6} to 10^{–3} *M*.

Ion chromatography can be used to determine chloride concentrations of 2–1000 ppb with a carbonate–bicarbonate eluent (23). Fluoride, nitrite, phosphate, bromide, nitrate, and sulfate do not interfere and can be measured simultaneously with a total analysis time of <30 min.

An ion-selective electrode is available for chloride analysis; chloride can be measured potentiometrically at 10^{–6}–1 *M*. Iodide and sulfide are the principal interferences.

Sulfate. Mine drainage may contribute to high sulfate concentrations resulting from pyrite oxidation. High concentrations may exhibit a cathartic action. Sulfate concentrations of 10^{–5} to 10^{–3} *M* can be titrated in an alcoholic solution with standard barium chloride and using Thorin as an indicator. Barium reacts with Thorin to form a deep red complex. The sample should be kept at pH 2–5, and the sample must be passed through a strong cation-exchange column prior to the titration. This is done to remove any multivalent cations which would also form intensely colored complexes with Thorin. High concentrations of sulfate can be determined by gravimetric analysis. The sulfate precipitates as barium sulfate. The ion-chromatographic response to sulfate is linear from 2 to 10⁴ ppb (23).

Nitrate and Nitrite. Nitrate is usually present in trace quantities in surface waters but occasionally occurs in high concentrations in some groundwaters. If present in excessive amounts, it can contribute to the illness infant methemoglobinemia. Nitrate is an essential nutrient for many photosynthetic autotrophs. Nitrite is an intermediate in the reduction of nitrate as well as in the oxidation of ammonia; it is also used as a corrosion inhibitor in some industrial processes.

Nitrite can be determined by reaction with sulfanilamide to form the diazo compound, which couples with *N*-(1-naphthyl)ethylenediamine dihydrochloride to form an intensely colored red azo dye. Nitrate can be determined in a similar manner after reduction to nitrite. Suitable reducing agents are cadmium filings or hydrazine. This method is useful at a nitrogen concentration of 10^{–7}–10^{–4} *M*.

Nitrate can also be measured potentiometrically with an ion-selective electrode at 10^{–5}–10^{–1} *M* (24,25). This method is suggested as a screening method for determining the approximate nitrate concentration (20). Ion chromatography can be used for nitrate concentrations of 2 to 10⁴ ppb (23).

Fluoride. A fluoride concentration of ca 1 mg/L is helpful in preventing dental caries. Fluoride is determined potentiometrically with an ion-selective electrode. A buffer solution of high total ionic strength is added to the solution to eliminate variations in sample ionic strength and to maintain the sample at pH 5–8, the optimum range for measurement. (Cyclohexylenedinitrilo)tetraacetic acid (CDTA) is usually added to the buffer solution to complex aluminum and thereby prevent its interference. If fluoroborate ion is present, the sample should be distilled from a concentrated sulfuric acid solution to hydrolyze the fluoroborate to free fluoride prior to the electrode measurement (26,27).

Several colorimetric procedures for fluoride are available, but it is usually desirable to distill the sample from concentrated sulfuric acid prior to analysis to eliminate interferences. One method is based upon bleaching a dye formed by the reaction of zirconium and sodium 2-(*p*-sulfophenylazo)-1,8-dihydroxy-3,6-naphthalenedisulfonate (SPADNS reagent) (28).

Phosphate. Phosphorus occurs in water primarily as a result of natural weathering, municipal sewage, and agricultural runoff. The most common form in water is the phosphate ion. A sample containing phosphate can react with ammonium molybdate to form molybdophosphoric acid (H₃P(Mo₃O₁₀)₄). This compound is reduced with stannous chloride in sulfuric acid to form a colored molybdenum-blue complex, which can be measured colorimetrically. Silica and arsenic are the chief interferences.

Some water samples contain phosphorus forms other than phosphate, eg, polyphosphate, hexametaphosphate, and organic phosphates. These forms can be hydrolyzed to phosphate in hot sulfuric acid solution and determined by the preceding method. The more refractory organic phosphates require digestion in a sulfuric acid–ammonium persulfate solution. Ion chromatography can also be used to measure PO₄^{3–} at 2 to 10⁴ ppb (21).

Boron and Borates. Boron is an essential element for plant growth; however, concentrations >2 mg/L are harmful to some plants. Natural-water concentrations of boron are usually well below this value, although higher concentrations can occur as a result of industrial waste effluents or cleaning agents.

Two colorimetric methods are recommended for boron analysis. One is the curcumin method, where the sample is acidified and evaporated after addition of curcumin reagent. A red product called rosocyanine remains; it is dissolved in 95 wt % ethanol and measured photometrically. Nitrate concentrations >20 mg/L interfere with this method. Another colorimetric method is based upon the reaction between boron and carminic acid in concentrated sulfuric acid to form a bluish-red or blue product. Boron concentrations can also be determined by atomic absorption spectroscopy with a nitrous oxide–acetylene flame or graphite furnace. Atomic emission with an argon plasma source can also be used for boron measurement.

Silica. The silica content of natural waters is usually 10^{–5} to (5 × 10^{–4}) *M*. Its presence is considered undesirable for some industrial purposes because of the formation of silica and silicate scales. The heteropoly-blue method is used for the measurement of silica. The sample reacts with ammonium molybdate at pH 1.2, and oxalic acid is added to reduce any molybdophosphoric acid produced. The yellow molybdosilicic acid is then reduced with 1-amino-2-naphthol-4-sulfonic acid and sodium sulfite to heteropoly blue. Color, turbidity, sulfide, and large amounts of iron are possible interferences. A digestion step involving NaHCO₃ can be used to convert any molybdate-unreactive silica to the reactive form. Silica can also be determined by atomic absorption with a nitrous oxide–acetylene flame or by atomic emission involving either a d-c or inductively coupled argon plasma source.

Oxygen. The solubility of atmospheric oxygen in water depends primarily on pressure, temperature, and salt content. The most reliable results for oxygen measurement are obtained from fresh samples. The longer the time lag between sampling and measurement, the greater the chance that the dissolved oxygen concentration diminishes because of chemical or biological activity in the sample. The Winkler titration has been the preferred method for

dissolved oxygen determination for many years. Several modifications of the basic method have been made to minimize interferences (20). The basis of this analysis is the quantitative oxidation of alkaline manganous hydroxide by the oxygen in the sample. Upon acidification in the presence of excess iodide, an amount of iodine equivalent to the dissolved oxygen is released. The iodine can then be titrated with standard sodium thiosulfate.

Considerable care is required in the collection of samples for dissolved-oxygen analysis. The errors associated with sampling can be minimized by an *in situ* method. Membrane-covered dissolved-oxygen electrodes are ideally suited to this purpose. These electrodes can be either polarographic or galvanic. The electrodes are protected by an oxygen-permeable polymeric membrane, which provides a rigorously defined diffusion barrier against solution impurities. For the polarographic type, an external voltage must be applied to the working or sensing electrode and must be sufficient to reduce molecular oxygen. With the galvanic type, two solid metal electrodes are used, such that the reduction of oxygen occurs spontaneously and no external voltage source is necessary. In both cases and under steady-state conditions, the resulting current is proportional to the dissolved-oxygen concentration in solution. The use and operation of dissolved-oxygen electrodes are thoroughly described in several publications (29,30).

Minor Mineral Constituents and Gases

Metals. The method of choice for metal analysis generally depends upon the concentration as well as the number of metals to be analyzed per sample and, to a lesser degree, the number of samples. When there is a large number of metals per sample, an emission spectroscopy method is preferred because of its simultaneous multielement capabilities. This technique has become much more popular for water analysis because of the development of a new emission source, the argon plasma. This source combines the advantages of the flame with the high temperature of an electric arc or spark. The plasma can be generated either through an induction process or by a direct current. When coupled with a polychromator in a direct-reading fashion, this method can be used to measure up to 60 elements in approximately 1 min (31–33). Limits of detection for many elements are comparable to or better than flame atomic absorption and, in some cases, to graphite-furnace atomic absorption.

Atomic absorption spectroscopy is more suited to samples where the number of metals is small, because it is essentially a single-element technique. The conventional air–acetylene flame is used for most metals; however, elements that form refractory compounds, eg, Al, Si, V, etc, require the hotter nitrous oxide–acetylene flame. The use of a graphite furnace provides detection limits much lower than either of the flames. A cold-vapor-generation technique combined with atomic absorption is considered the most suitable method for mercury analysis (34).

Nonmetals.

Arsenic. Total arsenic concentration can be determined by reduction of all forms to arsine (AsH₃) and collection of the arsine in a pyridine solution of silver diethyldithiocarbamate. Organoarsenides must be digested in acidic potassium persulfate prior to reduction. The complex that forms is deep red, and this color can be measured spectrophotometrically. Reduction is carried out in an acidic solution of KI–SnCl₂, and AsH₃ is generated by addition of zinc.

Atomic absorption spectroscopy is an alternative to the colorimetric method. Arsine is still generated but is purged into a heated open-end tube furnace or an argon–hydrogen flame for atomization of the arsenic and measurement. Arsenic can also be measured by direct sample injection into the graphite furnace. The detection limit with the air–acetylene flame is too high to be useful for most water analysis.

Selenium. Selenium is determined by atomic absorption after the organoselenides are broken down with acidic persulfate and all forms of selenium have been converted to H₂Se. The reduction is brought about in acidic solution of KI–SnCl₂ or borohydride, and H₂Se is generated by addition of zinc. The dihydrogen selenide is purged into an open-end tube furnace or argon–hydrogen flame for atomization and measurement. Selenium can also be determined by direct sample injection into the graphite furnace.

Cyanide. Industrial processes frequently discharge significant concentrations of cyanide, which can be extremely toxic at very low levels. Cyanide compounds are classified as either simple or complex. It is usually necessary to decompose complex cyanides by an acid reflux. The cyanide is then distilled into sodium hydroxide to remove compounds that would interfere in analysis. Extreme care should be taken during the distillation as toxic hydrogen cyanide is generated. The cyanide in the alkaline distillate can then be measured potentiometrically with an ion-selective electrode. Alternatively, the cyanide can be determined colorimetrically. It is converted to cyanogen chloride by reaction with chloramine-T at pH <8. The CNCl then reacts with a pyridine barbituric acid reagent to form a red-blue dye.

Bromide and Iodide. The spectrophotometric determination of trace bromide concentration is based on the bromide catalysis of iodine oxidation to iodate by permanganate in acidic solution. Iodide can also be measured spectrophotometrically by selective oxidation to iodine by potassium peroxymonosulfate (KHSO₅). The iodine reacts with colorless leucocrystal violet to produce the highly colored leucocrystal violet dye. Greater than 200 mg L of chloride interferes with the color development. Trace concentrations of iodide are determined by its ability to catalyze ceric ion reduction by arsenous acid. The reduction reaction is stopped at a specific time by the addition of ferrous ammonium sulfate. The ferrous ion is oxidized to ferric ion, which then reacts with thiocyanate to produce a deep red complex.

Both of these halides can also be determined potentiometrically with an appropriate ion-selective electrode. Sulfide and cyanide both interfere with the electrode response.

Gases.

Hydrogen Sulfide. Sulfide ion from 10^{–7} to 1 M can be measured potentiometrically with an ion-selective electrode. Mercuric ion interferes at concentrations >10^{–7} M. The concentration of hydrogen sulfide can be calculated knowing the sample pH and the pK_a for H₂S.

Ammonia. The most reliable results for ammonia are obtained from fresh samples. Storage of acidified samples at 4°C is the best way to minimize losses if prompt analysis is impossible. The sample acidity is neutralized prior to analysis. Ammonia concentrations of 10^{–6}–0.5 M can be determined potentiometrically with the gas-sensing, ion-selective electrode. Volatile amines are the only known interferents.

The most common colorimetric technique involves a reaction between ammonia and a reagent containing mercuric iodide in potassium iodide (Nessler reagent) to form a reddish-brown complex. Turbidity, color, and hardness are possible interferences that can be removed by preliminary distillation at pH 9.5.

Organic Materials

Nonspecific Organics.

Biochemical Oxygen Demand. The biochemical oxygen demand (BOD) test is an empirical determination of the oxygen requirement of a sample. It is most often applied to wastewaters, industrial effluents, and polluted waters. The decrease in the dissolved oxygen concentration resulting primarily from biological action is measured after storage for 5 d at 20°C.

If the sample has very low initial dissolved oxygen, it must be aerated; if it has a very high BOD, it may be necessary to dilute the sample. The dilution water is a phosphate buffer containing magnesium sulfate, calcium chloride, and ferric chloride. The percentage dilution must usually be determined by trial and error, although rough estimates may be made depending on sample type (20).

The dissolved oxygen concentrations are determined immediately and after five days. The method for dissolved measurement involves either a modified Winkler titration or a membrane-covered oxygen electrode. The difference between initial and final dissolved oxygen multiplied by the dilution factor is the BOD value.

Chemical Oxygen Demand. The chemical oxygen demand (COD) test measures the oxygen equivalent of the organic matter in a sample that is susceptible to oxidation by a strong oxidant (20). The sample is refluxed with a known amount of potassium dichromate in sulfuric acid for 2 hs. Silver sulfate is added to catalyze the oxidation of straight-chain compounds, and mercuric sulfate is added to the sample prior to refluxing to react with chloride ion and prevent its oxidation. The amount of unreacted dichromate is then titrated with standard ferrous ammonium sulfate.

Organic Carbon. The total organic carbon (TOC) in a water sample is determined by injecting a microliter sample into a heated, packed tube in a stream of oxygen. The water is vaporized and carbon is converted to carbon dioxide, which is detected with a nondispersive infrared analyzer. Nitrogen is bubbled through the acidified sample prior to injection to remove inorganic carbon and other volatiles.

Detergents. The most widely used surfactant in synthetic detergents is the readily biodegradable linear alkyl sulfonate (LAS). Since the

detergent industry began using this ingredient, the occurrence of foaming incidents has all but disappeared (see SURFACTANTS). The methylene-blue method is used to measure LAS in addition to alkyl sulfates and alkyl poly(ethoxyl sulfate)s. The blue salt that forms is extracted with chloroform and measured colorimetrically. There are many organic and inorganic interferences associated with this method and, since a large number of surfactants react with methylene blue, the test is designated as one for methylene-blue-active substances. If this test shows a significant amount of surfactant, confirmation of the LAS should be made by other methods, eg, thin-layer chromatography followed by infrared spectroscopy (35).

Oil and Grease. Industrial processes contribute most of the grease and oil found in water. Oil and grease can cause problems in wastewater-treatment processes as well as prevent the use of the sludge as fertilizer. The amount of oil and grease can be determined by extraction of an acidified sample with an organic solvent followed by evaporation of the solvent and weighing the residue. During the last part of the solvent evaporation, the sample must not be overheated to prevent the loss of low boiling oils.

Specific Organics. Large quantities of specific organic materials are used annually in industrial and agricultural applications, and these compounds or their degradation products are present in surface and groundwaters. The volume of publications dealing with analytical methodology for specific organics is enormous. The development of glass capillary columns has improved the application of gas-liquid chromatographic (glc) separations. In addition, the application of the mass spectrometer (ms) as a detector for gas-liquid chromatography has made the positive identification of peaks possible. High performance liquid chromatography (hplc), which involves various detectors, can be used to measure hydrophilic and hydrophobic organic compounds in water.

Various methods for the glc monitoring of EPA Consent Decree Priority Pollutants in water have been described (36) (see REGULATORY AGENCIES). The determination of organic pollutants in water by glc and ms methods has also been detailed (37,38). Nonvolatile organic compounds in drinking water have been determined by hplc (39) (see WATER, POLLUTION).

Pesticides. Chlorinated hydrocarbon insecticides are determined with an electron-capture detector following extraction with an organic solvent (20). If polychlorinated biphenyls (PCBs) are known to be present or if the extract contains so many pesticides that separation by glc is difficult, the extract should be passed through a Florisil column. Elution of the column with different solvents allows certain group separations of the pesticides. The following organochlorinated pesticides and PCBs have been determined by the method: lindane, heptachlor, heptachlor epoxide, aldrin, dieldrin, *p,p'*-1,1-dichloro-2,2-bis(*p*-chlorophenyl)ethane (*p,p'*-TDE), *p,p'* 1,1,1-trichloro-2,2-bis(*p*-chlorophenyl)ethane (*o,p'*-DDT), *o,p'*-DDT, *p,p'*-1,1-dichloro-2,2-bis(*p*-chlorophenyl)ethylene (*p,p'*-DDE), endrin, *p,p'*-methoxychlor, α -endosulfan, β -endosulfan, *cis*-chlordane, *trans*-chlordane, Aroclor 1248, Aroclor 1254, and Aroclor 1260. Quantitation is by comparison of chromatograms with standard concentrations of pure compounds treated in an identical manner. The phenoxy acid herbicides (2,4-dichlorophenoxy)acetic acid (2,4-D), silvex, and (2,4,5-trichlorophenoxy)acetic acid (2,4,5-T) can be determined by electron-capture detection after extraction and conversion to the methyl esters with BF₃-methanol. The water sample must be acidified to pH <2 prior to extraction with chloroform.

Identification of the pesticides is based on retention time on at least two dissimilar glc columns. A nonpolar and a relatively polar packing are generally used, for example, OV-17 and a mixture of QF-1 and DC-200.

Organophosphorus pesticides are determined with a flame photometric detector, which is sensitive only to compounds containing phosphorus or sulfur (20). The water sample is extracted with benzene, and the solvent is concentrated to a small volume prior to glc analysis. The following organophosphates are determined by this method: ethyl guthion, guthion, trithion, ruelene, diazinon, di-syston, ethion, imidan, malathion, methyl parathion, methyl trithion, parathion, thimet, and trolene (see INSECT CONTROL TECHNOLOGY).

High performance liquid chromatography with electrochemical detection has been used to determine 2–7 ppb of carbamate pesticides in water (40). The investigated pesticides were aminocarb, asulam, *sec*-butyl phenylmethylcarbamate (BPMC), carbaryl, carbendazim, chlorpropham, desmedipham, and phenmedipham.

Trihalomethanes. Wherever chlorine is used as a disinfectant in drinking-water treatment, trihalomethanes (THMs) generally are present in the finished water. The THMs usually formed are trichloromethane (chloroform), bromodichloromethane, dibromochloromethane, and tribromomethane (bromoform). There are four main techniques for the analysis of THMs: headspace, liquid-liquid extraction (lle), adsorption-elution (purge-trap), and direct aqueous injection. The final step in each technique involves separation by gas-liquid chromatography with a 2 mm ID coiled glass column containing 10 wt % squalene on chromosorb-W-AW (149–177 μ m (80–100 mesh)) with detection generally by electron capture.

The purge-trap method is the most widely accepted method for THMs as well as for other purgeable organohalides (41). The purgeable organics are stripped from the water sample with a stream of inert gas and are adsorbed in a porous polymer trap. The compounds are thermally desorbed from the trap into a gas chromatograph. All of the lle methods involve extraction of a small volume of water with a much smaller volume organic solvent (42). An aliquot of the extract is then gas-chromatographed. Different lle techniques have been evaluated to be as sensitive and accurate as the purge-trap methods (42,43).

Radioactive Materials

Radioactivity in environmental waters can originate from both natural and artificial sources. The natural or background radioactivity usually amounts to <100 mBq L. The development of the nuclear power industry as well as other industrial and medical uses of radioisotopes (qv) necessitates the determination of gross alpha and beta activity of some water samples. These measurements are relatively inexpensive and are useful for screening samples. The gross alpha or beta activity of an acidified sample is determined after an appropriate volume is evaporated to near dryness, transferred to a flat sample-mounting dish, and evaporated to dryness in an oven at 103–105°C. The amount of original sample taken depends on the amount of residue needed to provide measurable alpha or beta activity.

Alpha counting is done with an internal proportional counter or a scintillation counter. Beta counting is carried out with an internal or external proportional gas-flow chamber or an end-window Geiger-Mueller tube. The operating principles and descriptions of various counting instruments are available, as are techniques for determining various radioelements in aqueous solution (20,44). A laboratory manual of radiochemical procedures has been compiled for analysis of specific radionuclides in drinking water (45). Detector efficiency should be determined with commercially available sources of known activity.

Bacteria

A bacteriological examination of water is primarily carried out to determine the possible presence of harmful microorganisms. Testing is actually done to detect relatively harmless bacteria called *colon bacilli*, commonly called the coliform group, which are present in the intestinal tract of humans and animals. If these organisms are present in a water in sufficient number, then this is taken to be evidence that other harmful pathogenic bacteria may also be present.

One standard test used to determine the presence of the coliform group is called the multiple-tube fermentation technique (sometimes called the presumptive test). If this test indicates the presence of these bacteria, then a confirmed test must be done. If only negative colonies or no colonies develop during this test, it is considered negative; otherwise, a completed test must be undertaken. Positive results obtained in the completed test are evidence for the presence of coliform bacteria. Testing methods have been given by the APHA, and the detailed procedures contained therein should be consulted (20).

Another standard test, which is much simpler and more convenient, is the membrane filter technique. A suitable volume of sample is filtered through a sterile, 0.45- μ m membrane filter. The filter is placed in a petri dish containing a specific growth medium (M-Endo nutrient broth, M-Endo medium) and incubated for 24 h at 35°C. If after this time the colonies show the characteristic green sheen, this is taken as positive evidence for the presence of the coliform group (see WATER, SEWAGE).

Fecal Coliforms. Fecal coliforms are those originating from the intestines of warm-blooded animals. Fecal coliforms can be determined by a multiple-tube procedure, which must be applied to a positive presumptive test for optimum recovery of fecal coliforms (20). Incubation must be at 44.5 $\pm$ 0.2°C for 24 $\pm$ 2 h. Gas production during incubation is positive evidence of fecal coliform pollution.

A membrane filter technique can also be used to determine the presence of fecal coliforms, and this procedure is said to be 93% accurate (20). A sample is passed through a membrane filter, and this filter is placed in a petri dish containing an enriched lactose medium. The dishes are incubated at

44.5 ± 0.2°C for 24 h. Following the incubation period, the fecal coliform colonies appear blue.

Both multiple-tube and membrane-filter methods are also available for testing for the fecal streptococcal group (20). These assays can be used to provide supplementary data regarding the bacteriological quality of water. Other fecal indicators should also be used concurrently because of the survival characteristics of the fecal streptococci.

BIBLIOGRAPHY

"Water (Municipal)" in *ECT* 1st ed., Vol. 13, pp. 946–962, by H. O. Halvorson, University of Illinois; "Water (Analysis)" in *ECT* 2nd ed., Vol. 21, pp. 688–707, by Marvin W. Skougstad, U.S. Geological Survey; "Water (Analysis)" in *ECT* 3rd ed., Vol. 24, pp. 315–326, by R. B. Smart, West Virginia University and K. H. Mancy, University of Michigan.

1. F. C. Smith and R. C. Chang, *Crit. Rev. Anal. Chem.* **9**, 197 (1980).
2. J. D. Mulik and E. Sawicki, eds., *Ion Chromatographic Analysis of Environmental Pollutants*, Vol. 2, Ann Arbor Science Publishers, Inc., Ann Arbor, Mich., 1979.
3. W. B. Robbins and J. A. Caruso, *Anal. Chem.* **51**, 889A (1979).
4. B. V. L'vov, *Spectrochim. Acta Pt. B* **33B**, 153 (1973).
5. R. K. Skogerboe, *Toxicol. Environ. Chem. Rev.* **2**, 209 (1978).
6. L. T. Skeggs, *Am. J. Clin. Pathol.* **28**, 311 (1957).
7. J. Ruzicka and E. H. Hansen, *Anal. Chim. Acta* **78**, 145 (1975).
8. *Ibid.*, **99**, 37 (1976).
9. L. R. Snyder, *Anal. Chim. Acta* **144**, 3 (1980).
10. J. K. Foreman and P. B. Stockwell, eds., *Topics in Automatic Chemical Analysis*, Vol. 1, Ellis Horwood Ltd., W. Sussex, UK, 1979.
11. J. K. Foreman and P. B. Stockwell, eds., *Automatic Chemical Analysis*, Ellis Horwood Ltd., W. Sussex, UK, 1975.
12. T. M. Florence and G. E. Batley, *Crit. Rev. Anal. Chem.* **9**, 219 (1980).
13. M. J. Fishman, D. E. Erdmann, and T. R. Steinheimer, *Anal. Chem.* **53**, 182R (1981).
14. D. S. Polcyn, *J. Water Pollut. Control Fed.* **53**, 620 (1981).
15. F. B. DeWalle, D. Norman, J. Sung, E. S. K. Chian, and M. Giabbai, *J. Water Pollut. Control Fed.* **53**, 659 (1981).
16. C. E. Hamilton, ed., *Manual on Water*, 5th ed., ASTM Special Technical Pub. No. 422A, American Society for Testing and Materials, Philadelphia, Pa., 1978.
17. M. A. Forbes, ed., *Analytical Methods Manual*, Inland Waters Directorate, Environment Canada, Ottawa, Ontario, Can., 1979.
18. J. K. Kopp and G. D. McKee, eds., *Methods for Chemical Analysis of Water and Wastes*, 3rd ed., EPA 600/4-79-020, Washington, D.C., 1979.
19. H. S. Hertz, W. E. May, S. A. Wise, and S. N. Chesler, *Anal. Chem.* **50**, 428A (1978).
20. *Standard Methods for the Examination of Water and Wastewater*, 15th ed., American Public Health Association, Washington, D.C., 1981.
21. *Instrumentation for Environmental Monitoring*, rev., Vol. II, Water LBL-1, Technical Information Department, Lawrence Berkeley Laboratory, University of California, Berkeley, Calif., 1980.
22. H. Small, T. S. Stevens, and W. C. Bauman, *Anal. Chem.* **47**, 1801 (1974).
23. R. A. Wetzol, C. L. Anderson, H. Schleicher, and G. D. Crook, *Anal. Chem.* **51**, 1532 (1979).
24. D. Langmuir and R. I. Jacobson, *Environ. Sci. Technol.* **4**, 835 (1970).
25. D. R. Keeney, K. H. Byrnes, and J. J. Green, *Analyst* **95**, 383 (1970).
26. M. S. Frant and J. W. Ross, Jr., *Anal. Chem.* **40**, 1169 (1968).
27. J. E. Harwood, *Water Res.* **3**, 273 (1969).
28. E. Bellack and P. J. Schoubof, *Anal. Chem.* **30**, 2032 (1968).
29. K. H. Mancy and T. Jaffe, *Analysis of Dissolved Oxygen in Natural and Wastewater*, U.S. Public Health Service Publication No. 999-WP-37, Washington, D.C., 1966.
30. M. L. Hitchman, *Measurement of Dissolved Oxygen*, John Wiley & Sons, Inc., New York, 1978.
31. A. F. Ward, *Am. Lab.* **10**, 79 (1978).
32. R. M. Barnes, *Application of Plasma Emission Spectrochemistry*, Heyden & Sons, Inc., Philadelphia, Pa., 1979.
33. G. W. Johnson, *Spectrochim. Acta Pt. B*, **34B**, 197 (1978).
34. G. E. Millward and A. LeBihan, *Water Res.* **12**, 179 (1978).
35. H. H. Hellman, *Fresenius Z. Anal. Chem.* **293**, 359 (1978).
36. L. H. Keith, K. W. Lee, L. P. Provost, and D. L. Present, *ASTM Special Technical Publication STP 686*, American Society for Testing and Materials, Philadelphia, Pa., 1979, pp. 85–197.
37. D. Beggs, *NBS Spec. Publ. (US)* **519**, 169 (1979).
38. W. E. Pereira and B. A. Hughes, *J. Am. Water Works Assoc.* **72**, 220 (1980).
39. B. Crathorne, C. D. Watts, and M. Fielding, *J. Chromatogr.* **185**, 671 (1979).
40. J. L. Anderson and D. J. Chesney, *Anal. Chem.* **52**, 2156 (1980).
41. T. A. Bellar and J. J. Lichtenberg, *J. Am. Water Works Assoc.* **23**, 234 (1974).
42. R. C. Dressman, A. A. Stevens, J. Fair, and B. Smith, *J. Am. Water Works Assoc.* **71**, 392 (1979).
43. M. M. Varma, M. R. Siddique, K. T. Doty, and A. Machis, *J. Am. Water Works Assoc.* **71**, 389 (1979).
44. K. Haberer and U. Stuerzer, *Fresenius Z. Anal. Chem.* **299**, 177 (1979).
45. H. L. Krieger, *U.S. NTIS PB Report No. 253258*, National Technical Information Service, Washington, D.C., 1976, 61 pp.

General References

D. Midgley and K. Torrance, *Potentiometric Water Analysis*, John Wiley & Sons, Inc., Chichester, UK, 1991.

Standard Methods for the Examination of Water and Wastewater, 18th ed., American Public Health Association, Washington, D.C., 1992.

SUPPLY AND DESALINATION

Of the surface of the earth, 71% (3.60 × 10⁸ km²) is covered by oceans; their average depth is 6 km and their volume is 8.54 × 10⁸ km³. Unfortunately, this huge quantity of water is not suitable for very many human uses. Water with over 1000 ppm (parts per million by weight, or mg L) salt is usually considered unfit for human consumption, and water with over 500 ppm is considered undesirable, but in some parts of the world, people and land animals are forced to survive with much higher concentrations of salts, sometimes of over 2500 ppm.

Freshwater with less than 500 ppm (or 0.05%) dissolved solids is generally considered to be potable. Rain is the source of freshwater, and its precipitation of >1.3 × 10¹² m³ d over the earth's surface averages about 1.05 m (depth) per year. Extremes range from almost zero in North Chile's desert bordering the Pacific Coast to > 25.4 m in some tropical forests and on some high slopes where the high, cold mountains condense floods from the clouds.

Even rain is not pure water. Reports from the U.S. Geological Survey show that it contains 2.3–4.6 ppm of solids, or a yearly precipitation of 2.5–5 t km². Recently (ca 1997), work conducted in the United States and Europe has underscored the rather dangerous results of increased use of fossil fuels, where the SO_x and NO_x emissions that end up in the rain lower its pH from 5.6 (slightly acidic) for uncontaminated rain, to <pH4 for acid rains. Such acid rain has serious effects on surface waters (1). About 40 × 10⁶ t of SO_x and 25 × 10⁶ t of NO_x were emitted in the United States in 1980. There are, however, encouraging trends: the 1970 Clean Air Act has led to a gradual reduction in these emissions, bringing the SO_x emissions down from the previous levels cited by 10% by 1990, and the NO_x emissions down by 6%, with a consequent slight decrease in rain acidity. A part of the Clean Air Act is also intended to cap SO_x emissions from major point sources at 13.5 × 10⁶ t (2). Between 1994 and 1995, total SO_x emissions in the U.S. decreased remarkably by 13%; and total NO_x emissions by 8%.

About 60% of the land area of the earth is arid or semiarid and is not generally considered habitable. Mountainous areas and the polar areas covered by ice must be subtracted from the remainder to arrive at a land area that can be counted as available for human habitation and agriculture. Of this small fraction remaining, certain preferred land areas support the expanding world population more generously; hence, increasingly large amounts of the better lands are used for living, ie, for urban areas, industries, roads, airfields, etc. Areas for living and industry usually also require much more water per unit than do farm lands. Thus, as population and industry increase and the resulting demands for water multiply, agricultural areas with adequate water tend to become smaller.

The oceans hold about 97% of the earth's water. More than 2% of the total water and over 75% of the freshwater of the world is locked up as ice in the polar caps. Of the remaining 1% of total water that is both liquid and fresh, some is groundwater at depths of > 300 m and therefore impractical to obtain, and only the very small difference, possibly 0.06% of the total water of this planet, is available for human use as it cycles from sea to atmosphere to land to sea. Only recently have humans been able to regulate that cycle to their advantage, and even now (ca 1997), only infinitesimally, in some few isolated places.

Wells produce groundwater, stored from previous rains. However, the fact that in recent years wells have had to be made deeper and deeper to reach water shows that groundwater is being used faster than it is being replenished. Water lying in deep strata for millions of years is being mined like other minerals, never to be replaced. In Libya, oil drilling found a lake 100 m below the dry sand, hundreds of square kilometers in area, and 750 m deep (3). It has been estimated that this lake will supply irrigation of 800 km² for 300 yr, but the pumping of this water is as final an act as the pumping of Libya's petroleum, which probably dates from the same lush geological era. Once pumped, neither resource can be replenished.

Transport of Freshwater

For centuries, containers have been used to carry freshwater, usually for longer distances than would be practical for conduits. Trucks and railways have used tanks. Ships have carried and still do carry water halfway around the world in ballast tanks; tankers that otherwise would be returning empty from oil deliveries may make the return voyage filled with freshwater as precious to the oil-rich, water-poor country as the oil is to its market. Systems of dams, canals, and aqueducts were developed to carry freshwater considerable distances to growing cities, and to irrigate agricultural lands. Records of ancient Rome indicate that 14 aqueducts, averaging over 144 km in length, carried 1.175 × 10⁶ m³ d of water from the surrounding highlands by gravity. The Romans depended on gravity flow in open channels, apparently with little knowledge of predicting the friction losses in pipes, even though Pliny lists standard lead pipe in circumferences up to ca 2.5 m. The aqueducts of Istanbul are even more dramatic. The engineers' competence developed with increasing needs for water, and conduits became longer.

In more modern times, New York City's water system, initiated ca 135 yr ago, stands comparison as a true engineering marvel. Farsighted action in the late nineteenth century also gave the city extensive upstate watershed rights. This system, with a storage capacity of 2.07 × 10⁹ m³, safely furnishes about 8.5 million people with an average of 5.3 × 10⁶ m³ d of water. The system is not adequate today, but it has served well except in years of serious drought. Water enters the city via two tunnels, one built in 1917 and the other in 1937. The projected cost of a third tunnel was estimated at \$3.5 × 10⁹ in 1981 dollars, but could end up being as high as \$5–6 × 10⁹ by the time the tunnel is completed, with the first two stages currently scheduled for completion in 2010 (4). Increased population in the vicinity of the water supply reservoirs, a watershed approximately the size of the state of Delaware and as far as 160 km from the city, is increasingly threatening the water quality there. The city has therefore taken steps in the recent years to establish a comprehensive watershed protection program.

The 1974 Safe Drinking Water Act put all public water supplies in the United States under Federal supervision. For example, the 94th Congress authorized a six-state water study, the High Plains Study Council, one of several such studies conducted in past years, to develop a set of plans for increasing water supply in a region which supplies about 15% of the nation's wheat, corn, sorghum, and cotton, and 38% of its livestock, and which depends on its rapidly dwindling groundwater (the Ogallala aquifer), mindful of the Dust Bowl drought disaster of the 1930s (5). In 1990 it was estimated that the aquifer would be practically depleted by the year 2020. It was proposed that huge amounts of water from the Missouri and Arkansas Rivers be diverted to the High Plains, with an estimate by the Army Corps of Engineers that it would take as long as 9 yr to design, 25 yr to build, and would cost \$6–25 × 10⁹ in 1982 dollars. Rising energy costs, costs of construction, local objections to Federal intervention, environmental and navigation considerations, and various other problems have prevented this and many other large water supply programs from materializing.

The problem of bringing water to southern California continues to be one producing controversy. The region is essentially a desert, devoid of any significant water resources, yet its population is growing at a fast pace: in 1990 it stood at 30 million and is expected to reach 49 million in 2020, with half of the increase expected to occur in the arid South Coast region (6). Water is transported from northern California and from the Colorado River, at distances of up to about 1000 km, expected to increase to about 1600 km by the year 2000, often against the objections by those living around these sources. The project to build the so-called Peripheral Canal, a 67-km long, 120-m wide channel to carry water from the Sacramento River delta to an existing aqueduct and then to the agriculturally productive San Joaquin Valley and thence to Southern California, was passed by California's legislature in 1980, and then defeated in 1982 owing to political opposition from a coalition of environmental and local urban and agricultural interests. Elevated salinity in the California Aqueduct and some other water canals has recently revived some interest in this project.

In some places and under certain conditions, freshwater can be obtained more cheaply by desalination of seawater than by transporting water. This is true when all the costs of extremely large monetary investments in dams, reservoirs, conduits, and pumps to move the water are considered. Before the rapid escalation of fuel costs between 1973 and 1980, the cost of desalination of seawater to adequately supply southern California would have been less than that of transport to the Peripheral Canal. This would have been the case even if there were an unlimited supply of water in the mountains of northern California, a condition that does not appear to exist. It has been shown that before 1973 a seacoast town could have been supplied with $7\text{--}12 \times 10^4 \text{ m}^3 \text{ d}$ of freshwater more cheaply by desalination than by damming and piping water a distance of $>160 \text{ km}$ (7). Indeed, the 1987–1992 drought in California has compelled the city of Santa Barbara to construct a water desalination plant, and a $76,000\text{-m}^3 \text{ d}$ plant is planned for the western coast of Florida (8).

Because construction costs have remained approximately constant since the mid-1980s, despite inflation, desalination has a high potential of becoming less costly with improved technology, whereas the cost of transporting water is not likely to decrease substantially. Desalination increases the total amount of freshwater; transportation never does. The reduction of water costs, if possible at all, will, however, be limited. Desalination is an energy-intensive process, and no presently conceivable alternatives promise any reduction of energy costs. The cost savings obtained as a consequence of advances in desalination technology have been counterbalanced by the increase in energy costs since the oil embargo of 1973.

The Water Problem

Many cities of the world do not levy a separate fee on water distributed, and even in those places where water is in shortest supply, a minimal ration may be free to everyone. The problem of wasted water and unmetered water adding to the overall water demand is not new. In ancient Rome, fountains were connected to the public water by privately installed and owned lead pipelines, many of which were unrecorded, illegal, and hence untaxed. Frontius, the water commissioner of Emperor Nerva of Rome in AD 96, developed crude meters to increase revenue and cut demand.

Today in the United States, three times as much water is used per capita than in 1900; with inclusion of all industrial and agricultural uses, this quantity is probably ten times as great as at the turn of the century. Individual usage in some southern cities, with swimming pools, lawns, air conditioners, and other local demands, can be as much as twice the national average. Population increase multiplies the total withdrawals, particularly in cities, where they may be as much as $1.06 \text{ m}^3 \text{ d}$ person. In New York City, for example, which did not practice adequate water metering, per capita consumption grew from $0.69 \text{ m}^3 \text{ d}$ person in 1970 to $0.76 \text{ m}^3 \text{ d}$ person in 1981. Improved metering and water conservation have reversed this trend over the recent years, and in 1995 the per capita consumption fell to $0.68 \text{ m}^3 \text{ d}$ person, about the same as in 1970 (4). Philadelphia is also able to meter or bill only a part of the supplied water, and the total per capita water consumption has grown from 1970 to 1990 by 30%, to about $0.9 \text{ m}^3 \text{ d}$ person, whereas the metered and billed accounts are expected to have a consumption of only $0.35 \text{ m}^3 \text{ d}$ person (9).

The use of ground and surface freshwater in the United States from 1950–1990 is shown in Figure 1 (10). Whereas the overall increase in use had been slightly reduced and then arrested after 1980, this was primarily because of a reduction in use for irrigation and industry. In some part it is a result of conservation and improved efficiency, but in large part it is the result of a decline in agricultural and industrial production. Public supply continues to increase, approximately keeping pace with population growth.

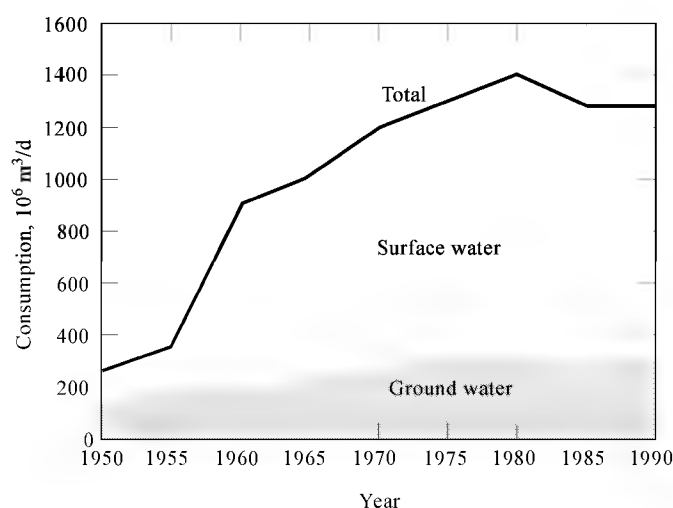


Fig. 1. Trends of estimated ground and surface freshwater use in the United States, 1950–1990 (10).

An analysis of freshwater use in the United States predicted that by the year 2040, barring major changes in use patterns and natural water cycles, withdrawals and consumption will increase by about 40% (11). Most of this increase must be met by reuse, and thus water recycling should be raised significantly. For example, in manufacturing, recycling currently provides a reuse ratio, defined as the quantity ratio of (water actually used) (its natural water source), of 1.3, which must be increased by the year 2020 to above 6.3. The balance between water supply and demand is worsening most seriously in the Rio Grande, upper and lower Colorado, Great Basin, High Plains, and California water resources regions of the United States.

In some areas of the United States, twice as much water is pumped from the ground as soaks into it. Although the amount of groundwater within 0.8 km of the surface is estimated at $3.8 \times 10^{16} \text{ m}^3$, in some places the water table has dropped by 1–5 m for each year of the present generation, thus exhausting a historical treasure. Not only is the groundwater being depleted, but often this withdrawal causes sinking of the ground level, as it has near Houston, Texas; Mexico City; and in Florida. Las Vegas, Nevada, is growing rapidly upon a ground level that has sunk about a meter in recent years as a result of greatly increased mining of prehistoric water. This water supplies many acres of swimming pools, many thousands of "tons of refrigeration" for air conditioning, and other water uses. It is water from wells in a desert where only 7.5–10 cm of rain falls each year. Over-pumping since the 1920s, mostly to supply the needs of Southern California, has caused the land in the California San Joaquin Valley to sink by as much as 9 m.

Southern California is suffering from seawater intrusion in certain aquifers because of overpumping. To minimize the problem, the Orange County Water District in 1975 began a pioneering effort, known as Water Factory 21, to treat $57,000 \text{ m}^3 \text{ d}$ of sewage by secondary effluent treatment processes, purify about one third of the product water by reverse osmosis (RO), blend the remainder of the treatment product with the RO process product and with high quality groundwater, and inject the mixture into the aquifer. Apart from recharging thereby the aquifer, it also reduces, if not stops, seawater intrusion into it (12,13) (see REVERSE OSMOSIS). The plant continues to produce drinking-quality water for injection with no adverse effects on groundwater.

Water is far from evenly distributed in the United States, with major shortages in some very populous areas. California and the Southwest have always been water-short, but in the first half of the 1960s and the late 1970s, the northeast, too, experienced water shortages. Particularly ominous was the recent six-year (1987–1992) drought in California in which the water supplies were insufficient to meet the demands, an event that portends even worse shortages in the future, of 8.6×10^9 – $11.1 \times 10^9 \text{ m}^3 \text{ yr}$ in drought seasons, and 4.6×10^9 – $7.0 \times 10^9 \text{ m}^3 \text{ yr}$ in normal seasons, by the year 2020 (6). Restrictions of water use in many states have become virtually annual occurrences in years of low rainfall. Whereas the overall water supply in the United States is expected to meet demand in the foreseeable future, serious imbalances are expected to continue and worsen owing to geographic, seasonal, and annual variations in the supplies (11).

Two out of five U.S. cities have inadequate water supplies, and at least a quarter of the U.S. population faces serious water shortages. Yet, half the

states, having two-thirds of the industry and over half of the population, have direct access to as much as they can draw of the approximately $3.3 \times 10^8 \text{ km}^3$ of seawater. The solids content of this water, mainly salts, varies worldwide from 25,000 ppm (2.5‰) in the Baltic Sea to over 45,000 ppm (4.5‰) in some of the more confined gulfs of the Indian Ocean. The waters of the wide oceans are almost constant, at 35,000 ppm (3.5‰). Also, many inland areas have access to large quantities of water too brackish to drink.

Increasing pollution of water sources worldwide is a major contributor to the climbing shortage of usable water. A recent report of the U.S. Environmental Protection Agency (14) points out that about 40% of the U.S. rivers, lakes, and estuaries surveyed are too polluted for even basic uses such as fishing or swimming. Especially grim is the condition of the Great Lakes, which contain one-fifth of the world's fresh surface water. About 97% of the Great Lakes waters were found to be substandard for designated uses, and particularly worrisome is the fact that the lakes are continuously polluted by toxic chemicals. Some of the good water in the U.S. is considered to be "threatened," and a projection of need for sewage treatment costing over \$100 billion was made. In this context it is important to note the vital importance of water-quality legislation. Along with the Clean Air Act, the Clean Water Act of 1972 is of unprecedented value in U.S. environmental protection history. Launched to "restore and maintain the chemical, physical, and biological integrity of the Nation's waters," it has resulted in a significant slowdown of the pollution processes and in many cases reversed the trend. Many rivers and creeks, for example, such as the Delaware and Schuylkill Rivers passing through Philadelphia, which had owing to severe pollution lost most aquatic life and became inaccessible for recreation, are now largely restored again. The Delaware River, for example, became septic in the 1940s and 1950s, dissolved oxygen concentration dropped to zero in the warmer months, and fish became extinct. The situation now, while still not perfect, is highly improved, with herring and shad swimming upstream to spawn. The activities in support of the Clean Water act incur, however, an economic cost, and the Act is continuously threatened by interest groups which object to this cost and inconvenience. Many countries in the world, especially the developing ones, and those currently or formerly under communist regimes, still do not have such legislation or enforcement and have severely polluted waters. A World Bank report issued in 1995 stated that 80 countries with 40% of the world's population already have water shortages which could cripple agriculture and industry, and that about 95% of the world's urban areas dump raw sewage into rivers and other water bodies, with dirty water killing 10 million or more people every year and causing untold economical damage. The water supply and quality situation is worsening rapidly. It was estimated that the developing countries alone would need $\$600\text{--}800 \times 10^9$ for water projects over the next 10 years, with no identifiable available funding sources.

Some of the most attractive areas of the world, particularly islands and beaches, are almost devoid of freshwater. The biggest and one of the fastest growing of the industries of the world is tourism. It is a particularly attractive industry to developing countries, and in some of these it may be almost the only nonagricultural industry. Tourism accounts for a substantial use of the available freshwater, and it may be stifled or entirely prevented in otherwise attractive places if there is insufficient freshwater. To conserve available water supplies, some hotels might use double water systems, ie, seawater for flushing, but they still must provide on the order of 400–600 L d per tourist to assure a comfort to guests unaccustomed to water shortages. Production of these quantities of water by desalination techniques has become an important expense to the hotels. Today, the cost of production of desalinated water can be as high as $\$4.00\text{--}5.25 \text{ m}^3$, and in some locations even higher.

The island of Bermuda, for example, derives the largest percentage of its income of any country from tourism, receiving over 500,000 visitors per year. It has no surface water and its ground water became increasingly saline as a result of excessive pumping. By law, each house there must have no less than four-fifths of its roof area guttered for catching rainwater, and a covered tank of 100 imperial gallons (0.45 m³) capacity for every 10 ft² (0.9 m²) of prescribed area of catchment to store the rainwater. It is estimated that about 4000 m³ d is acquired in this manner, serving as the main potable water source for the island (15). The U.S. Virgin Islands has similar regulations. Vegetation is cleared from hill slopes above the natural limestone and coral, which are plastered to give drains and catchments for heavy seasonal rains. But even with all this, some water had to be imported until adequate desalination capacity was introduced, at this time totaling 30,000 m³ d, with a water cost of $\$4.76 \text{ m}^3$ (16).

The newly acquired wealth of the countries in the Middle East has given an unexpected boost in desalination for the region. Starting with nearly zero, Saudi Arabia, Kuwait, and the United Arab Emirates have over the past three decades added plants producing about $9 \times 10^6 \text{ m}^3 \text{ d}$ of freshwater, representing nearly one-half the worldwide desalination capacity (17). This includes the world's largest single-site, 946,000-m³ multistage flash distillation (MSF) complex, in Al-Jubail, Saudi Arabia; the world's largest (2,400-m³ d each) MSF units in Abu Dhabi; and the world's largest (113,600 m³ d) reverse osmosis seawater desalination plant, in Jedda, Saudi Arabia. Following the Arab Peninsula in order of desalination capacity are the United States, North Africa, the rest of the Middle East, the European Mediterranean area, Southeast Asia, the rest of Europe, and the Caribbean islands. Water desalination is increasingly used for the treatment of effluent waters for reuse, and of river and city waters to improve purity for various industrial applications such as boiler feed and ultrapure water for the electronics industry.

In many of the developing nations, lack of water hampers the profitable exploitation of material resources (18). For example, proven mines on Egypt's Red Sea Coast cannot be operated; others on the same coast are operable only with desalination water for processing phosphate ores and for mining lead and zinc. The workers were rationed 15–35 L d of water. Fishing industries on South America's arid Pacific Coast cannot be expanded for lack of water to process the haul. These represent great losses in the world's supplies of minerals and foods; there are many others. Not all freshwater shortages are in the torrid tropics; a substantial iron-ore deposit in a small waterless island off the coast of Iceland needs water for workers. Needed is a desalination plant or some means of utilizing as liquid water the heavy fogs that prevail there. Offshore oil-drilling rigs around the world similarly need water. Technological progress in the last decade has made readily available packaged small (10–400 m³ d) seawater RO and vapor-compression systems. In fact, virtually all offshore drilling rigs today have their own seawater conversion plants. Desalination facilities, and in turn desalinated water, become readily available, contingent only on the availability of funds.

Cities on the lower reaches of a large river, with many cities above, use water that has been through sewers upstream many times. On the lower Mississippi, a water inventory indicates such reuse averages 14 times, with biochemical oxidation of the wastes during the flow between cities. The Rhine River, in passing through several countries, all of which drink from and dump into its waters, is subject to international problems of pollution. Of considerable import here is the fact that many so-called hard contaminants present in sewage increase almost proportionally with reuse of the water. Such materials do not ferment or oxidize under ordinary sewage treatments, and some are known carcinogens, which presents serious health problems to potable waters drawn from these bodies.

The U.S. government spent ca $\$150 \times 10^6$ in fiscal 1969 on water resources research relating to artificial rainmaking, soil conservation, waste treatment, desalting, public health, and planning research. This level of annual expenditure, large compared to the rest of the world, has in the meantime diminished (19), even though it was even then considered to be rather small for supporting a need estimated at $>\$100 \times 10^9$ for water facilities worldwide within 10–15 yr (20). Several government agencies conduct water research, primarily the U.S. Geological Survey, the Environmental Protection Agency, the Bureau of Reclamation, and the Department of Agriculture. Practically all of the reports and results of research on water in the United States are available from the National Technical Information Service, U.S. Department of Commerce, Springfield, VA 22161.

Saline Water for Municipal Distribution. Only a very small amount of potable water is actually taken by people or animals internally, and it is quite uneconomical to desalinate all municipally piped water, although all distributed water must be clear and free of harmful bacteria. Most of the water piped to cities and industry is used for little more than to carry off small amounts of waste materials or waste heat. In many locations, seawater can be used for most of this service. If chlorination is required, it can be accomplished by direct electrolysis of the dissolved salt (21). Arrayed against the obvious advantage of economy, there are several disadvantages: use of seawater requires different detergents; sewage treatment plants must be modified; the usual metal pipes, pumps, condensers, coolers, meters, and other equipment corrode more readily; chlorination could cause environmental pollution; and dual water systems must be built and maintained.

Pipes, valves, fittings, and almost all other components of small equipment are now available in plastic or ceramics, which do not corrode in salt water and are less expensive than the metals now used. Synthetic detergents are now available for use with seawater, although a final rinse with freshwater may be desired. Saltwater sewage can be treated successfully. Dual water systems using freshwater and seawater are already in use on ships and in many island resort hotels. Many of these also have seawater systems for fire fighting. This trend will grow.

Some inland municipalities now distribute water with salt content exceeding 1000 ppm, water so brackish as to be unpleasant to the taste, even though it is distributed as potable water. Each home may produce or purchase the very small requirement of freshwater for drinking and cooking. Small

membrane and ion-exchange desalinators are available; these produce the relatively few liters of potable water required (see MEMBRANE TECHNOLOGY; ION EXCHANGE). They can be purchased from and periodically serviced by service companies in many communities. The cost per liter of potable water produced in the home is several times as great as the cost in a central desalination plant, but the amount of desalinated water needed is only a small fraction of the total supply.

Home desalinators are possible only for industrialized countries with a central service organization. They will eventually become available on a rental service contract basis, as is standard practice for water softeners in many communities. Although rental of water softeners is common in the United States, home membrane-system rental is not established.

Alternatively, small amounts of potable water may be delivered by truck to distribution centers or to tanks on house roofs. This system exists in Kuwait, which has many filling stations from which tank-truck operators buy water at \$1.00 m³ for distribution at about \$3.00 m³. Although much water is directly piped to residences in Kuwait today, 12% of the people still get their water by truck. In Khartoum, Sudan, families that buy from vendors, who deliver sacks of water by donkey, pay an average of \$16 per month.

Water in Industry. Freshwater for industry can often be replaced by saline or brackish water, usually after sedimentation, filtration, and chlorination (electrical or chemical), or other treatments (22). Such treatment is not necessary for the largest user of water, the electric power industry, which in the United States passed through its heat exchangers in 1990 about 40% of the total supply of surface water, a quantity similar to that used for agriculture, and it was 48% of the combined fresh and saline water withdrawals (10). Single stations of 1000 MW may heat as much as 12 Mm³ d by as much as 10–15°C.

The power-plant cooling water is either returned directly to its source, such as rivers, lakes, or oceans, or is recycled by circulation and consequent cooling in cooling towers. Cooling towers circulate the warmed water downward against a rising stream of air which removes heat by evaporating a part of the water to cool the balance. In the United States, cooling towers are not so common as in Europe because of the large bodies of water available.

Construction of new power plants in the coal region of the western United States presents serious problems in states whose laws dictate zero effluent. In these plants, cooling-tower water withdrawn from rivers cannot be returned to them. In these situations, cooling-tower effluent is purified by distillation (vapor-compression plants have predominated) and by a combination of distillation and membrane technology. The converted water then is used as boiler feedwater; the plant blowdown (effluent) is evaporated from open-air lined pools, and pool sediment is periodically buried back in the coal mine with the flue ashes.

Although 600 m³ of water is used to make a metric ton of fertilizer, 150–240 m³ to make a tonne of steel, 480 m³ to make a tonne of gasoline, and 1000 m³ to make a tonne of acetate fiber, little if any is required chemically in any of these processes. Recycling can reduce industrial requirements by a factor of 10–50. Much of this water, particularly that for cooling, and often that for washing, can be saline. Some petroleum refiners have used salt water to remove heat (water's principal role in gasoline production), and some have actually produced table salt by evaporation in cooling towers.

The pulp and paper industry has tried for many years to use salt water for some of the 250–400 m³ of water required to make a metric ton of paper (see PULP; PAPER). Here, however, salt is disadvantageous to the chemical processes, either in pulping the lignocellulose or in the recovery of values from the black liquor after pulping, and can corrode expensive papermaking machinery. The possibility of recovering and reusing at least part of this water after membrane processing is under study.

The textile industry, also a large consumer of water, must always be located in areas with abundant water supply (see TEXTILES—SURVEY). Before long, advances in purification of textile industry effluents are expected to free the industry of this limitation and enable textile plants to locate in virtually any area. The great fluid reservoir of heat, the alternative to water, fresh or salt, is the atmosphere. Increasingly, industry is using air coolers to dissipate the large quantities of heat from power plants, petroleum distillation, and other process use. This would reduce the use of cooling water correspondingly.

Air coolers almost invariably add considerably to plant cost, but they are competitive in operating cost based on direct once-through use of water that requires no treatment. If the alternative to air coolers is the use of water that requires substantial treatment or pumping costs, the air coolers will cost less to operate.

Water for Agriculture. Two liters of water in some form is the daily requirement of the average human, depending on many personal and external conditions. However, at least several hundred liters per day are required for the growing of vegetables, fruits, and grain that make up the absolute minimal daily food ration for a vegetarian. About 0.48 m³ is required to produce an egg, and 31 m³ to produce a kilogram of beef, based on the cereals required by the animals. The quality of water (in particular, its salinity) is of considerable importance to agriculture. There are crops that can grow in water of relatively high salinity (for example, sugar beets tolerate salinity of up to 5,000 ppm), but most crops cannot tolerate salinity exceeding 1000 ppm. This limit led to the treaty between the United States and Mexico in which the United States agreed to limit the salinity of the lower Colorado River reaching Mexico to approximately 1000 ppm, a requirement that had to be met by constructing a 276,000 m³ d RO desalination plant (the largest RO plant constructed to date) in Yuma, Arizona, which would reduce the salinity of the agricultural drainage water from 3200 ppm to 250 ppm. The desalted water is then mixed with Colorado River waters to reduce their overall salinity to the required limits (23). Originally planned to start in 1981, various delays made it ready for starting only in 1983, a year in which unusually large water flow in the Colorado River made desalination unnecessary. The plant ran successfully in 1991–1992, but subsequent funding problems restricted capacity to only about one-third of the total, and then stopped operation altogether, but the plant is available for restarting if needed. The salinity problem is being resolved in the interim by diverting the saline waters to the marshy area of the Santa Clara Slough in Mexico.

A loaf of bread contains little of the more than a ton of water necessary to grow the wheat therein. The water content of vegetables comes from the 1000–2000 times their weight of it that is needed to grow them. Some of the main losses in agriculture may be reduced by agronomists and plant physiologists. Increasingly, the balance sheet for irrigation plant technology now makes use of agronomic WUE (water-use efficiency), which is a ratio of the amount of harvestable or economic biomass to the water consumed by evapotranspiration. Studies are being made to determine whether it is possible to supply less water to the roots, with better absorption there and smaller losses by transpiration through the leaves. For example, research includes tissue-culture work with protoplast fusion and recombinant DNA technologies to reduce stresses of salinity, drought, flooding, ion toxicities, nutrient deficiencies, temperature extremes, and photosynthetic efficiency. Certainly, plant structures of an entirely different type will be necessary to prevent, for example, the loss by transpiration in a field of corn on an August day, of an amount that may be equivalent to more than a centimeter of rainfall. However, the recently developed high yield grains, which use much less water than the conventional varieties, are a great aid to the developing nations.

It is possible to breed plants that have more efficient systems for utilization of water, and agricultural technology can help existing crop plants by spraying impervious coatings on them. Extremely small amounts of long-chain, fatty alcohols reduce evaporation losses from quiet lakes or reservoirs to less than 5% of the normal surface evaporation.

Waterproofing sandy soils to prevent drain-through has been successful in increasing crops as much as 400% with the same rainfall. A special plow lifts the soil to allow melted asphalt to be layered in overlapping impermeable strips 82 cm wide and 50 cm below the surface (18). Waterproofing the surface in Israel by compacting with chemicals increases runoff to basins or fields on slopes below. In many places, barren slopes have been coated with asphalt or concrete in layers as thin as 0.3 cm to catch rain, which is conducted to catch basins for irrigation or other uses. In St. Thomas, U.S. Virgin Islands, such catchment basins have, however, now been abandoned in favor of desalination.

The one-seventh of the world's crop lands that are irrigated produce one-quarter of the world's crops. Irrigation's main losses result directly from seepage and evaporation from the open water-carrying channels and the soil. Only a small fraction of the water withdrawn from the irrigation ditch or pipe is absorbed by the plants. Plastic films, as ground covers through which the plants protrude, prevent some losses but at great expense for film and labor. Cheaper systems are necessary to assure better water utilization by plants. Other possible goals would be food plants with membranes capable of separating freshwater from brackish water, to give a non-salty crop. Progress has been made in both of these directions, and some plants have been developed that accumulate salt from the ground.

Vegetables and fruits such as melons are profitably grown in four or five crops per year by hydroponics, ie, the growth of plants in large, shallow concrete tanks containing no soil but gravel and water with added nutrients. Such installations can even be found in areas known for their water scarcity such as Aruba and St. Croix (U.S. Virgin Islands) where both desalinated seawater and brackish water are used for the purpose. Much water is still necessary per unit weight of crops, but the largest losses of ordinary irrigation are prevented, as indeed they must be because of the comparatively high cost

of the water. Such concentrated agriculture is very expensive in preparation of land area, but economical for water and labor requirements. Production is high in the tropics, and hydroponics offers a major opportunity to many developing countries.

Other locations where desalinated water of either brackish or sea origin is used for agriculture are the Channel Island Guernsey, Israel, Libya, and countries in the Arabian Gulf including Saudi Arabia. The cost of such agricultural use is at best marginal and at worst exorbitant, depending on the amounts of water required.

Economic Aspects

As shown in Table 1, residential water rates in the U.S. have gone up significantly between 1978 and 1994 (24).

Table 1. Residential Water Costs, U.S. \$/1000 ft³ (28.3 m³)^a

Year	City					
	El Paso, Texas	Albuquerque, New Mexico	Los Angeles, California	Newark, New Jersey	Boston, Massachusetts	Philadelphia, Pennsylvania
1978	3.99	4.48	4.55	7.79	8.90	8.13
1994	7.89	10.54	15.80	15.13	19.92	14.07

^a Ref. 24.

According to the 1994 *National Water and Wastewater Rate Survey* (24), the lowest rate per 28.3 m³ of water in the United States was \$5.40 in Saginaw, Michigan, and the highest, \$44.88 in Scranton, Pennsylvania. Since the rates for wastewater were higher, by an average of \$2.59, increased rapidly from a surfeit of \$0.05 in 1986, and since an increasing number of utilities also charge customers for stormwater runoff treatment, the actual costs to the customer are higher than those shown in Table 1.

The range of consumer prices for water in the developed countries in 1990 ranged from about \$9.90 in Norway to \$45.28 in Greece, per 28.3 m³, or \$0.35–1.60 m³. In comparison, 1993 water costs from large desalination plants were in the \$1.50–2.00 m³ range (25).

Paradoxically, even rapid lowering of demand may cause rapid cost escalation; this was the case in New Jersey during a 1981 drought when conservation caused substantial shrinking of demand and private water companies had to double their water prices. The continued increase of demand and reduction of supply portend real and relentless water-cost increases in every part of the United States in the future. One possible way to assure at least adequate supplies, and possibly to moderate these cost increases, is through water reuse.

Desalination Development Programs and Associations

U.S. Department of the Interior: U.S. Geological Survey and the Bureau of Reclamation. The U.S. Department of the Interior (USDI), long active in many areas of water resources and management, was authorized by Congress in 1952 to organize the Office of Saline Water (OSW), later transformed into the Office of Water Research and Technology (OWRT), for research, development, and demonstration (RD&D) to develop economical processes for desalination. In 1996 dollars, the expenditure for this purpose rose from \$1.2 × 10⁶ in 1953, to a peak \$130.4 × 10⁶ in 1967, remaining around the \$100 × 10⁶ level through 1973. One fundamental principle guiding these efforts and enunciated by one President of the United States is that all scientific and engineering knowledge and skills developed in the wide reaches of this program are to be made immediately and freely available to every nation in the world, as a generous contribution to the betterment of the lives of all humankind.

This dissemination of knowledge has occurred, and desalination technology developed in the United States has resulted in a significant benefit to the world at large, even though the U.S. desalination industry has not always shared in the profits from the exploitation of this technology. In its effort, OSW and its successor, OWRT, supported the full range of activities, from fundamental through demonstration plants, covering all desalination technologies. In membrane technology, it was responsible for a number of landmark discoveries, including the desalination properties of cellulose acetate RO membranes, seawater conversion with membranes, construction of spiral-wound modules, composite-film seawater membranes, and very large membrane modules. The early demonstration distillation plants utilizing diverse distillation techniques have also made their mark (see HOLLOW-FIBER MEMBRANES). The results were published in more than 1,200 government reports, all in the public domain, and thousands of professional papers.

The training of many of today's industrial leaders in desalination technology was supported by these agencies.

Declaring that water desalination was a mature industry which did not require substantial governmental support any longer, the federal government stopped this RD&D desalination initiative in 1972, gradually reduced the annual expenditure for desalination to about \$2 × 10⁶ and closed the OWRT in 1982, with funding for this purpose continuing at this annual level and diminishing somewhat, under the auspices of the U.S. Geological Survey (USGS) and the Bureau of Reclamation (BuRec).

United Nations. The U.N. Department of Economic and Social Development; U.N. Children's Fund (UNICEF); U.N. Development Program; U.N. Environment Program; the U.N. University; the Economic Commissions for Africa, Europe, Latin America, and the Caribbean; the Economic and Social Commissions for Asia and the Pacific and Western Asia; the U.N. Centre for Human Settlements; the U.N. Disaster Relief Coordinator Office; the International Research and Training Institute for the Advancement of Women; and the World Food Program are all engaged in addressing water problems. The International Labor Organization, the Food and Agriculture Organization of the U.N., UNESCO, the World Health Organization, the World Bank, the World Meteorological Organization, United Nations Industrial Development Organization, and the International Atomic Energy Agency are also involved. The wide variety of water-related activities conducted by these many organizations is coordinated by the U.N. Intersecretariat Group for Water Resources, and a summary of them is given in Reference 26.

A milestone event was the United Nations Water Conference held in Mar del Plata, Argentina, in 1977, which dealt with the integrated planning and development of water resources, reported in the Mar del Plata Action Plan adopted by the Conference (27). The Action Plan contains a set of resolutions and recommendations on a wide spectrum of activities in the field of water resources development, covering such questions as assessment of water resources; water use and efficiency, including its development and use for sectoral purposes; environment, health, and pollution control; policy, planning, administration, and institutional aspects; education, training and research; natural hazards and regional and international cooperation. A major outcome of the conference was the General Assembly Resolution 35 18, which proclaimed the period 1981–1990 the International Drinking Water Supply and Sanitation Decade, committing member countries of the U.N. to a substantial improvement in the standards and levels of services in drinking water by 1990.

In 1990 the members of the Intersecretariat Group for Water Resources undertook regional assessments of the situation concerning key areas of the Mar del Plata Action Plan, in order to formulate a strategy for its implementation in the decade of the 1990s and beyond (28). Further stock-taking and planning was done at the International Conference on Water and the Environment—Development Issues for the 21st Century, held January 26–31, 1992, in Dublin, Ireland (26).

Sound international agreements on the sharing and proper maintenance of water resources are an important condition for peaceful coexistence among nations which have geohydrological links. The World Bank's vice president for Environmentally Sustained Development was quoted in 1995 to say that "many of the wars in this century were about oil, but wars of the next century will be over water." A good example is the negotiations on the future of water resources coordination in the course of the peace-building process between Israel and its Arab neighbors, in which the United Nations obviously has a role (29).

Details about the various U.N. water-related activities can be found in References 26–29, as well as in topical reports published by the United Nations.

International Desalination Association. The IDA (formerly IDEA) is the major association specifically devoted to desalination. Beginning in 1976 it began organizing international congresses in desalination and water reuse. The proceedings of its conferences have been published (30). It also publishes a trade magazine (31).

American Desalination Association. The ADA, formerly the National Water Supply Improvement Association (NWSIA), is devoted to water supply improvement and considers desalination as a major technique. The association publishes a newsletter, *ADA News*.

The European Desalination Society, formed recently, is the successor of the European Federation of Chemical Engineers Working Party on Desalination and Water Technology. The latter organization has held seven symposia in different parts of the world, and their published proceedings contain much valuable material (32).

Desalination: Manufactured Freshwater

Desalination has been used for providing drinking water on seafaring ships since ancient times (using solar or fuel heat), and an early reference to the scientific or miraculous conversion by Moses of bitter groundwater to fresh, viz, "... and the Lord showed him a wood and he put it into the water and the water became sweet," is made in the Old Testament (33). The possibility of producing freshwater from seawater or brackish water by separation of the salts opens a new dimension in the supply of freshwater. Areas bordering the sea would have an available raw material without limit or cost of transportation to the water facility. The successful realization of desalination by the combined effort of chemists, and chemical, mechanical, and materials engineers, as opposed to the search for and transport of existing freshwater, gives new hope for adequate water in many, but not all, cases. Figure 2 shows the remarkable growth of the cumulative contracted capacity of large desalination plants since 1959, when large-scale, land-based desalination began. As of the end of 1994, a total of 10,300 desalting plants having a capacity of at least $100 \text{ m}^3/\text{d}$ each, with a total capacity of $19,200,000 \text{ m}^3/\text{d}$, have been installed or contracted worldwide (17). This output can easily satisfy the domestic and public (excluding industrial and irrigation) water needs of at least 50 million people. About two-thirds of these plants are for desalting seawater, and one-quarter for desalting brackish water. About 70% of the desalted water is for municipal use, 20% for industrial use, and the remainder is used in the power industry, in discharge treatment, and by the military. There are more than 1700 companies and institutions involved in water desalination (34).

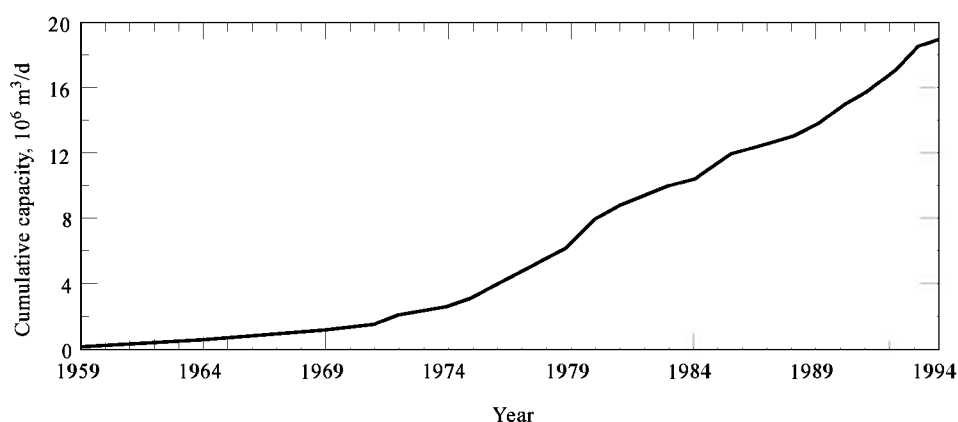


Fig. 2. History of the cumulative water production capacity of all existing and contracted land-based desalting plants which can each produce at least $100 \text{ m}^3/\text{d}$ of freshwater (17).

Cost remains the main stumbling block to desalination procedures, but freshwater is now one of those materials produced as it is needed. Water has significant differences from other manufactured materials, eg, the very low value of the commodity and the very large amount of the product needed. Thus, plant costs might be expected to be high compared to unit sales value of production, and the capital charges for the production equipment do make up a very large percentage of costs in comparison to other production industries.

Obstacles attend this new solution of the freshwater problem, which magnify those familiar to the chemical engineer in purifying other cheap or worthless raw materials into valuable products by treatment with chemicals or thermal or electrical energy. These obstacles are quite different from the previous main problem of water supply, ie, the factor of happenstance in finding a river or lake nearby or of making a fortunate geological strike.

In many places, the need for desalination is even more urgent than the production of food, which is limited by water shortages. These shortages exist both in the petroleum-rich countries and in many of the least-developed and poorest countries of the world. In the case of the former, the improved water supply obtained through desalination has already had stunning effects; by contrast, little hope for progress is on the horizon for the poor countries.

Water desalination is a process that separates the water from a saline water solution. The natural water cycle is the prevalent, and best, example of water desalination. Ocean waters evaporate as a result of solar heating and atmospheric influences, the vapor, consisting mostly of freshwater because of the negligible volatility of the salts at these temperatures, rises buoyantly and condenses into clouds in the cooler atmospheric regions, is transported across the sky by cloud motion, and is eventually deposited back on the earth's surface as freshwater rain, snow, and hail. The global freshwater supply from this natural cycle is ample, but, as mentioned before, many regions on earth do not receive an adequate share. Population growth, rapidly increasing demand for freshwater, and increasing contamination of the available natural freshwater resources render water desalination increasingly attractive.

In addition to freshwater, seawater is also a source for sodium, magnesium, chlorides, iodine, bromine, and magnesium metal (see **SODIUM COMPOUNDS**; **MAGNESIUM COMPOUNDS**; **IODINE**; **BROMINE**; **MAGNESIUM AND MAGNESIUM ALLOYS**). Many other elements are certain to be economically obtained from the ocean as technology for the recovery improves.

Many ways are available for separating water from a saline water solution. The oldest and still prevalent desalination process is distillation. The evaporation of the solution is effected by the addition of heat or by lowering of its vapor pressure, and condensation of these vapors on a cold surface produces freshwater. The three dominant distillation processes are multistage flash (MSF), multieffect (ME), and vapor compression (VC). Until the early 1980s, the MSF process was prevalent for desalination. Now membrane processes, especially reverse osmosis (RO), are economical enough to have gained about one-third of the market. In all membrane processes, separation occurs as a result of the selective nature of a membrane's permeability, which, under the influence of an external driving force, permits the preferential passage of either water or salt ions, but not both. The force driving the process may be pressure (as in RO), electric potential (as in electrodialysis, ED), or heat (as in membrane distillation, MD). A process used for low salinity solutions is the well-known Ion exchange (IE), in which salt ions are preferentially adsorbed onto a material that has the required selective adsorption property, and thus reduce the salinity of the water in the solution.

The cost of desalted water is composed of the capital cost of the plant, the cost of the energy needed for the process, and the costs of operation, maintenance staff, and supplies. In large seawater desalination plants the cost of water is about $\$1.4\text{--}2 \text{ m}^3$, dropping to less than $\$1 \text{ m}^3$ for desalting brackish water. A methodology for assessing the economic viability of desalination in comparison with other water supply methods is described in Reference 34. Desalination plants are relatively simple to operate, and progress toward advanced controls and automation is gradually reducing operator expenses.

Minimal Energy Requirements. The relative effect of the cost of the energy on the cost of the freshwater produced depends on local conditions, and is up to one-half of the total. In attempting to reduce this cost, it is of interest to determine the minimal energy amount thermodynamically needed for separating the water from the saline solution. The physical background to this will be introduced in a simple example. Because of the negligible

volatility of the salts, the vapor of an aqueous saline solution is practically pure H₂O. It is well known that the vapor pressure of a saline water solution at a constant temperature decreases as the solution salinity increases. Thus, the vapor pressure of such a solution is lower than the vapor pressure of pure water at the same temperature. Withdrawal of this pure water (here in vapor form) from the saline solution to a pure water product storage vessel which is maintained at the same temperature thus requires pumping of the vapor from the vapor under lower pressure above the saline solution to the vapor under higher pressure above the pure water. The work needed for that vapor pumping, assuming a 100^o efficient pump (compressor) and perfectly insulated vessels, is the minimal energy needed for separating the water from the saline solution.

The effect of salt concentration on the vapor pressure of the solution has its well-known temperature exposition, called the boiling point elevation. Because the vapor pressure of saline solutions is lower than that of pure water, their boiling point temperature is higher. The difference between the boiling points is called the boiling point elevation, which rises with the concentration, and it can thus be seen as the measure of the extra energy needed for separation, ie, here the required raising of the temperature of the solution by this amount to attain boiling. Thermodynamically reversible separation defines the minimal energy requirement for that process. The minimal energy of separation $W_{\min}$ in such a process is the change in the Gibbs Free Energy between the beginning and end of the process, ΔG. The minimal work when the number of moles of the solution changes from n_1 to n_2 is thus

$$W_{\min} = \int_{n_1}^{n_2} (\Delta G) \, dn_W$$

(1)

Bromley and co-workers (36) have calculated the minimal energy of separation of water from seawater containing 3.45 wt % salt, at 25°C, to be 2.55 kJ / (kg fresh water) for the case of zero fresh water recovery (infinitesimal concentration change) and 2.91 kJ / (kg fresh water) for the case of 25% fresh water recovery. $W_{\min}$ is, however, severalfold smaller than the energy necessary for water desalination in practice.

Improved energy economy can be obtained when desalination plants are integrated with power generation plants (37,38). If the power generation part of such a dual-purpose plant is of the Rankine (steam) type, a back-pressure or extraction turbine is typically used to supply the low pressure and temperature steam as the heat source for the desalination plant. If the power plant is of the gas-turbine type, the turbine exhaust gas is typically used as the heat source for a boiler which provides heating steam for the desalination plant. Diesel engine reject heat, both from the engine coolant and the exhaust can also be used as the heat source for desalination. Dual-purpose plants lead to important energy savings, but also to increases in capital cost and complexity of operation. It is important to be able to deal effectively with possible transient mismatches in power and water demand. Practically all of the operating distillation-type desalination plants with a capacity above 1000 m³ / d are of the dual-purpose type.

Materials and Scaling Issues. Two aspects of the basically simple desalination process require special attention. One is the high corrosivity of seawater, especially pronounced in the higher temperature distillation processes, which requires the use of corrosion-resistant, and therefore expensive, materials. Typical materials in use are copper–nickel alloys, stainless steel, titanium, and, at lower temperatures, fiber-reinforced polymers and special concrete compositions (39). It is noteworthy that in quest of a lower initial cost, the use of inadequate materials of construction in many locations combined with poor operation by virtually untrained hands led to rapid deterioration and failure of plants long before their estimated design life. Adequate experience suggests by now how to avoid such failures. The other aspect is scale formation (40,41), discussed in more detail below.

Obtaining maximum performance from a seawater distillation unit requires minimizing the detrimental effects of *scale* formation. The term *scale* describes deposits of calcium carbonate, magnesium hydroxide, or calcium sulfate that can form in the brine heater and the heat-recovery condensers. The carbonates and the hydroxide are conventionally called alkaline scales, and the sulfate, nonalkaline scale. The presence of bicarbonate, carbonate, and hydroxide ions, the total concentration of which is referred to as the alkalinity of the seawater, leads to the alkaline scale formation. In seawater, the bicarbonate ions decompose to carbonate and hydroxide ions, giving most of the alkalinity.



The kinetics of the formation of the magnesium hydroxide and calcium carbonate are functions of the concentration of the bicarbonate ions, the temperature, and the rate of release of CO₂ from the solution. At temperatures up to 82°C, CaCO₃ predominates, but as the temperature exceeds 93°C, Mg(OH)₂ becomes the principal scale. Thus, in seawater, there is a considerable tendency for surfaces to scale with an increase in temperature.

The interrelationship of nonalkaline scales (CaSO₄, CaSO₄·1/2 H₂O, CaSO₄·2H₂O) depends on temperature and the concentration of CaSO₄. To assure that no hemihydrate scale forms, MSF operators must run their plants in such a manner as to assure that the concentration of the total dissolved solids does not exceed 70,000 ppm at temperatures of 120°C. With average-salinity seawater, plants can operate at a concentration factor of 2, but in the Middle East where water salinity can be as high as 50,000 ppm, the concentration factor should not exceed 1.4. Under no circumstances should the total dissolved solids exceed 70,000 ppm, ie, twice the concentration of normal seawater at 120°C.

A number of options for controlling scale formation are used in plant operations around the world. One approach is use of mechanical means, including thermal shock. Although rare today (ca 1997), this practice can be found in use with the few obsolete submerged tube evaporators.

Another approach, the so-called seeding technique, provides preferential sites for the nucleation of scale, which permits the heat-transfer surfaces to remain clean of scale. Extensive studies of this technique have been conducted, and field use was reported in the former USSR as early as the mid-1960s (42). The use of ion-exchange methods is another possible approach. For calcium, the exchange can be represented as



Magnesium can be exchanged in a similar fashion. To date, cost considerations have prevented the use of ion exchange for scale control.

Another approach is the use of polyphosphate-based blends including proprietary chemicals. The exact mechanism of the observed effect is not completely understood. In the polyphosphate mode of operation, the polyphosphate is dosed in quantities of 2–5 ppm; periodically, sludge resulting from phosphate treatments is removed by acid cleaning (see DISPERSANTS).

In another option, the hydrogen ion from added acid decomposes the bicarbonate ions.



About 120 ppm of sulfuric acid must be provided for normal seawater. Control of acid dosing is critical; the amount of acid must be stoichiometric to the alkalinity expressed as CaCO₃. In conjunction with acid dosing, the CO₂ formed must be removed and some sodium hydroxide added to maintain ca pH 8 in the system. Alternatively, less than-stoichiometric amounts of acid can be added to retain some alkalinity in the untreated feed; in either case, CO₂ removal is done with packed columns. Acid-dosed feed is passed through a column with air flow that sweeps the CO₂ from the feed saturated with carbon dioxide. This is usually followed by a deaeration, during which both the air and CO₂ are reduced to the levels needed to minimize, if not eliminate, corrosion. Although acid-dosing does permit higher operating temperatures, it often has had a devastating effect on plant life.

The option most promising today involves use of a new family of polymers, the so-called high temperature scale-control chemicals. These are compounds that, added in 3–8 ppm, lead to lattice distortion and the formation of a nonadhering scale. Belgard (CIBA-GEIGY) was the first compound

exemplifying this type of MSF operation, which is now steadily displacing acid in the operation of MSF plants around the world, with important contribution to plant life (43) (see DISPERSANTS).

While the ambient-temperature operation of membrane processes reduces scaling, membranes are much more susceptible not only to minute amounts of scaling or even dirt, but also to the presence of certain salts and other compounds that reduce their ability to separate salt from water. To reduce corrosion, scaling, and other problems, the water to be desalted is pretreated. The pretreatment consists of filtration, and may include removal of air (deaeration), removal of CO₂ (decarbonation), and selective removal of scale-forming salts (softening). It also includes the addition of chemicals that allow operation without scale deposition, or which retard scale deposition or cause the precipitation of scale which does not adhere to solid surfaces, and that prevent foam formation during the desalination process.

Saline waters, including seawater, contain, besides a variety of inorganic salts, also organic materials and various particles. They differ in composition from site to site, and also change with time as a result of both natural and human causes. Design and operation of desalination plants requires good knowledge of the saline water composition and properties (41,44).

Desalination Plant Security. Ingestion into the plant of saline water feed which has been contaminated by undesirable components may not only impair product water quality, but also impair or incapacitate the plant for future operation. Membrane-based desalination plants are exceptionally sensitive to such damage. One of the obvious contaminants, especially in the oil-producing countries, is spilled oil. Heavy metals, detergents and other undesirable components which increasingly pollute the seas and oceans must be kept out of the plant intakes. Screens, filters, and oil booms are commonly used for protecting the plant intake. The latter have been especially appreciated during the Gulf War, in which the Iraqi forces intentionally dumped oil into the Gulf in order to incapacitate the desalination plants of the Gulf countries. Contamination also arose from their firing of the Kuwaiti oil wells and destruction of the water treatment facilities which released large amounts of sewage into the Gulf. As described in (43,44) they have also engaged in massive destruction of the Kuwaiti desalination and power plants.

Because of the vulnerability of desalination plants to damage which may arise from poor regional management, accidents, war, and terrorism, and the consequent severe impact on the water supply of countries which primarily depend on desalination, it is vitally important (1) to design the plant with robust safeguards against ingestion of undesirably contaminated saline water, (2) to ensure that regional resources' management prevents such contamination, (3) to provide for adequate freshwater storage, and (4) to provide adequate plant security.

The Major Desalination Processes

The major water desalination processes that are currently in use or in advanced research stages are described herein. Information on detailed modeling can be found in the literature cited. The major texts on water desalination written since the 1980s are those by Spiegler and Laird (47), Khan (48), which contains many practical design aspects, Lior (49) on the measurements and control aspects, Heitman (40) on pretreatment and chemistry aspects, and Spiegler and El-Sayed (50), an overview primer. Extensive data sources are provided in References 39 and 51.

Distillation Processes

Multistage Flash Evaporation (MSF). Almost all of the large desalination plants use the MSF process shown schematically in Figure 3. A photograph of a modern operating plant is shown in Figure 4. The seawater feed is preheated by internal heat recovery from condensing water vapor during passage through a series of stages, and then heated to its top temperature by steam generated by an external heat source. The hot seawater then flows as a horizontal free-surface stream through a series of stages, created by vertical walls which separate the vapor space of each stage from the others. These walls allow the vapor space of each stage to be maintained at a different pressure, which is gradually decreased along the flow path as a result of the gradually decreasing temperature in the condenser seawater-preheater installed above the free stream. The seawater is superheated by a few °C relative to the vapor pressure in each stage it enters, and consequently evaporates in each stage along its flow path. The latent heat of the evaporation is supplied by equivalent reduction of the sensible heat of the evaporating water, resulting in a gradual lowering of the stream temperature. The evaporation is vigorous, causing intensive bubble generation and growth with accompanying stream turbulence, a process known as flash evaporation (52–54). One of the primary advantages of the MSF process is the fact that evaporation occurs from the saline water stream and not, as in other distillation processes such as submerged tube and multiple-effect evaporation, on heated surfaces, where evaporation typically causes scale deposition and thus gradual impairment of heat-transfer rates. Also, the fact that the sensible heat of water is much smaller than its latent heat of evaporation, the specific heat $c_p = 4.182 \text{ kJ kg per } ^\circ\text{C}$ change of water temperature vs $h_{ig} = 2378 \text{ kJ kg}$, respectively, and that the top temperature is limited by considerations of scaling and corrosion, dictate the requirement for a very large flow rate of the evaporating stream. For example, (in the following subscripts b , d and s refer to distillate, brine, and steam, respectively) in operating between a typical top temperature $T_{b,t}$ of 90°C at the inlet to the evaporator and an exit temperature $T_{b,e}$ of 40°C corresponding to the ambient conditions, the overall temperature drop of the evaporating stream is 50°C. Using these values, the heat balance between the sensible heat of the water stream, flowing at a mass flow rate $\dot{m}_b$, and the latent heat needed for generating water vapor (distillate) at a mass flow rate $\dot{m}_d$ is

$$(\dot{m}_b - \dot{m}_d) c_p (T_{b,t} - T_{b,e}) \approx \dot{m}_d h_{fg}$$
(5)

which yields the brine-to-product mass flow ratio as

$$\frac{\dot{m}_b}{\dot{m}_d} = \frac{h_{fg}}{c_p (T_{b,t} - T_{b,e})} + 1 = \frac{2378}{(4.182) (50)} + 1 = 12.37$$
(6)

Therefore, 12.37 kg saline water are needed in this case to produce 1 kg distillate. This high flow rate incurs corresponding pumping equipment and energy expenses, sluggish system dynamics, and, because the stream level depth is limited to about 0.3–0.5 m for best evaporation rates, also requires large evaporator vessels with their associated expense.

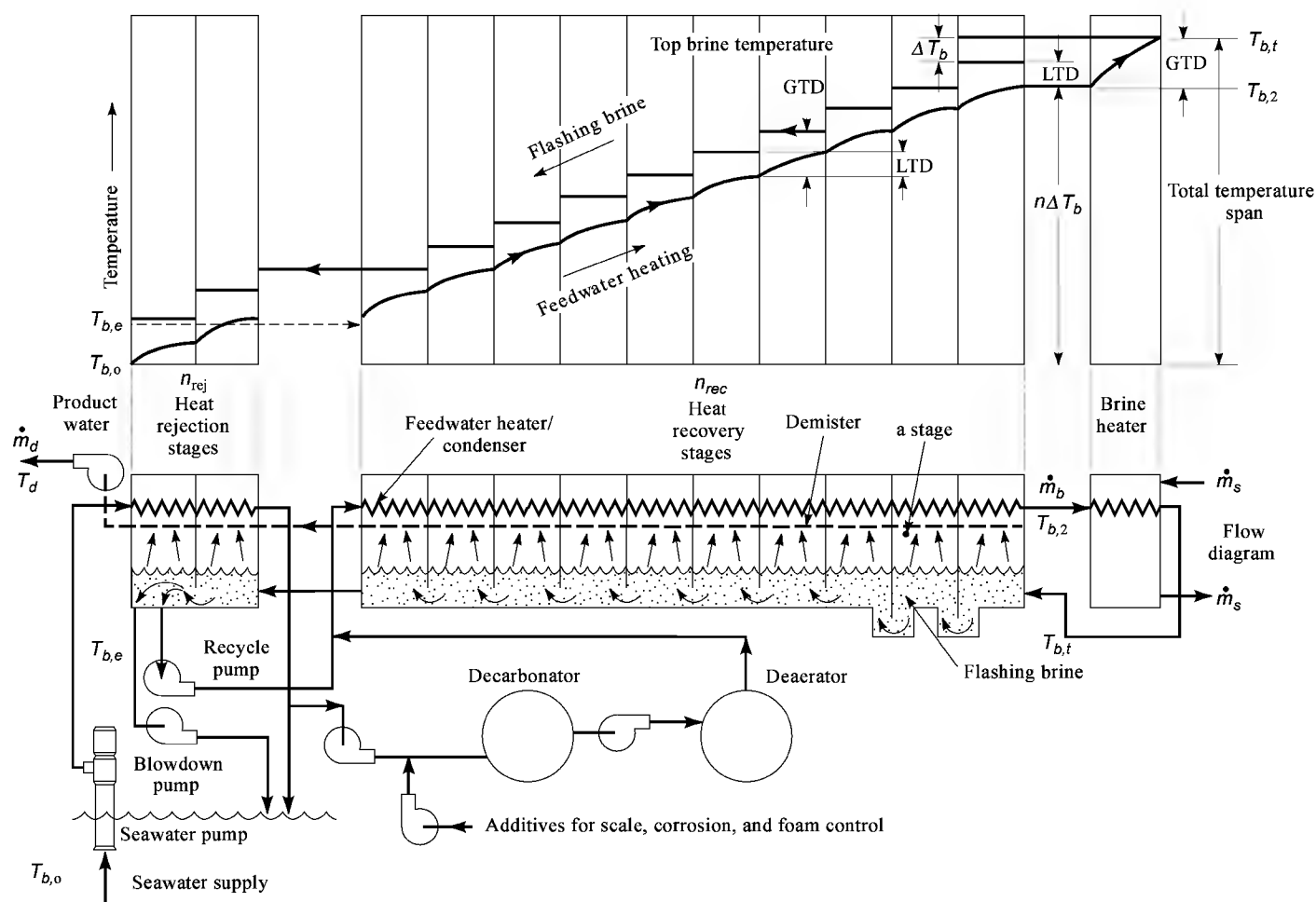


Fig. 3. Schematic flow and temperature diagram of the multistage flash (MSF) process for a recirculation type plant.

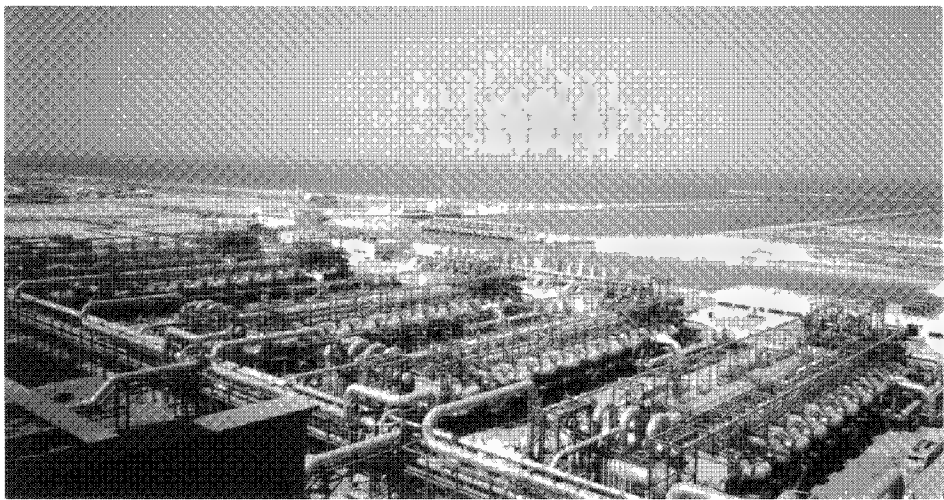


Fig. 4. The 341,000-m³ d multistage flash (MSF) evaporation desalination plant Al Taweelah B in Abu Dhabi, United Arab Emirates. Courtesy of Italmimpianti SpA. It is a dual-purpose plant, composed of six identical power and desalination units. The desalination units at 56,800 m³ d each are currently (1997) the largest in the world. They have 17 recovery and 3 reject stages and a Performance Ratio of 8:1. The plant also produces 732 MWe of power.

The generated water vapor rises through a screen (demister) placed to remove entrained saline water droplets. Rising further, it then condenses on the condenser tube bank, and internal heat recovery is achieved by transferring its heat of condensation to the seawater feed that is thus being preheated. This internal heat recovery is another of the primary advantages of the MSF process. The energy performance of distillation plants is often evaluated by the performance ratio, *PR*, typically defined as

$$PR \equiv \frac{\dot{m}_d}{\dot{m}_s} \tag{7}$$

where $\dot{m}_s$ is the mass flow rate of heating steam. Since the latent heat of evaporation is almost the same for the distillate and the heating steam, *PR* is also the ratio of the heat energy needed for producing one unit mass of product (distillate) to the external heat actually used for that purpose. Most of the heating of the brine stream to the top temperature $T_{b,t}$ is by internal heat recovery, and as seen in Figure 3, the external heat input is only the amount of heat needed to elevate the temperature of the preheated brine from its exit from the hottest stage at $T_{b,2}$ to $T_{b,t}$.

Assuming for simplification that the temperature drop of the flashing brine, ΔT_b , is the same in each stage and that the specific and latent heat of the brine remains the same throughout the plant, the relationship between the number of stages, *n*, and the performance ratio is expressed by

$$PR = \frac{1}{\frac{LTD}{T_{b,t} - T_{b,e}} + \frac{1}{n}}$$

(8)

where LTD is the lowest temperature difference between the flashed vapor and the heated feedwater, in each stage (Fig. 3). Equation 8 shows that increasing the number of stages increases the PR . This implies that more heat is then recovered internally, which would thus require a larger condenser brine-preheater heat-transfer area. The required heat transfer area, A , per unit mass of distillate produced for the entire heat recovery section (composed of n_{rec} stages), and taking average values of the overall vapor-to-feedwater heat transfer coefficient U and $LMTD$ per stage, is

$$A = \frac{h_{b,fg}}{U(LMTD)}$$

(9)

where $h_{b,fg}$ is the average latent heat of evaporation of the flashing brine, $LMTD$, the log-mean temperature difference between the vapor condensing on the tubes and the heated brine flowing inside the tubes, for an average stage defined as

$$LMTD = \frac{GTD}{\ln \frac{GTD}{LTD}} = \frac{(T_{b,t} - T_{b,2})}{\ln \left(\frac{T_{b,t} - T_{b,2}}{LTD} \right)} LTD$$

(10)

where GTD is the greatest temperature difference between the flashing brine and the brine heated in the condenser. The size of the heat transfer area per unit mass of distillate produced by the plant is

$$A = \frac{h_{fg,b}}{U} \frac{n_{rec}}{(T_{b,t} - T_{b,e})} \ln \left(\frac{n_{rec}}{n_{rec} - PR} \right)$$

(11)

Examination of this equation will show that the required heat transfer area for the heat recovery section per unit mass of distillate produced, A , increases significantly when PR is increased, and decreases slightly as the number of heat recovery stage, n_{rec} , is increased.

The MSF plant shown in Figure 3 is of the recirculation type, where not all of the brine stream emerging from the last evaporation stage is discharged from the plant, as it would have been in a once-through type of plant. A fraction of the emerging brine is mixed with pretreated seawater and recirculated into the condenser of the heat-recovery section of the plant. Because only a fraction of the entire stream in this configuration is new seawater, which needs to be pretreated (removal of air and CO_2 , ie, deaeration and decarbonation, and the addition of chemicals that reduce scale deposition, corrosion and foaming), the overall process cost is reduced. The recirculation plant is also easier to control than the once-through type.

Whereas most of the energy exchange in the plant is internal, steady-state operation requires that energy in an amount equal to all external energy input also be discharged from the plant. Consequently, the heat supplied in the brine heater plus any pumping energy is discharged in the heat-rejection stages section of the plant (Fig. 3). Assuming, for purposes of estimation, an equal temperature drop in each stage and that the pumping energy can be neglected relative to the heat input in the brine heater indicates that the ratio of the number of the heat-recovery to heat-rejection stages is approximately equal to the performance ratio PR .

Further detail about MSF desalination can be found in References 48, 55, and 56. A detailed design of an MSF plant producing 9,500 m³ d gallons of fresh water per day was published by the U.S. government (57).

Multi-Effect Distillation (ME). The principle of the multi-effect (ME) distillation process is that the latent heat of condensation of the vapor generated in one effect is used to generate vapor in the next effect, thus obtaining internal heat recovery and good energy efficiency. Such plants have been used for many years in the salt, sugar, and other process industries. Several ME plant configurations, most prominently the horizontal tube multi-effect (HTME) (Fig. 5) and the vertical tube evaporator (VTE), shown schematically in Figure 6, are in use. In the HTME, vapor is circulated through a horizontal tube bundle, which is subjected to an external spray of somewhat colder saline water. The vapor flowing in these spray-cooled tubes condenses, and the latent heat of condensation is transferred through the tube wall to the saline water spray striking the exterior of the tube, causing it to evaporate. The vapor generated thereby flows into the tubes in the next effect, and the process is repeated from effect to effect. A low temperature multi-effect distillation plant developed to operate at an upper limit of 70°C, illustrated in Figure 5, has a low fuel consumption, estimated by the manufacturer at 2.4–2.8 kg fuel per 1000 m³, and lower capital cost because the whole plant is built of aluminum (58). Many such plants have been installed and operated successfully.

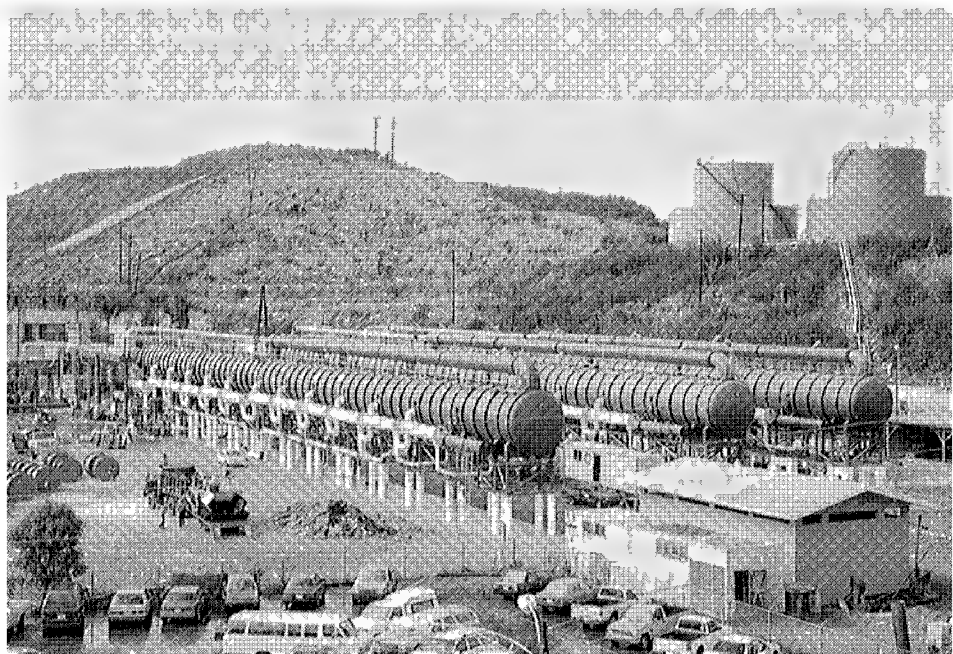


Fig. 5. A horizontal-tube multi-effect (HTME) desalination unit, producing 5000 m³ d in St. Croix, U.S. Virgin Islands. Courtesy of I.D.E. Technologies Ltd.

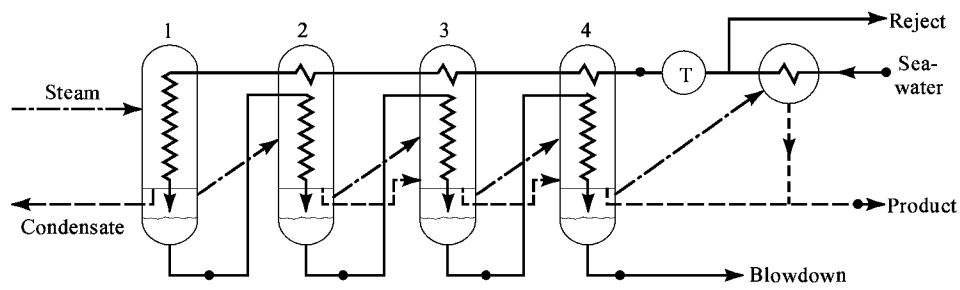


Fig. 6. Simplified schematic flow diagram of a typical 4-effect vertical-tube multi-effect (VTE) desalination plant, where (—) represents brine, (●) represents vapor, () represents condensate, ① denotes pretreatment, elements enclosed by a cylinder-like shape constitute an effect, and ●s represent pumps.

In the vertical tube multi-effect evaporator (VTE), the saline water typically flows downward inside vertical tubes, and evaporates as a result of condensation of vapor coming from a higher temperature effect, on the tube exterior. While internal heat recovery is a feature common to both MSF and ME processes, there are at least three important differences between them. One is that evaporation in the ME process occurs on the heat-transfer surfaces (tubes), whereas in the MSF process it takes place in the free stream. This makes the ME process much more susceptible to scale formation. At the same time, the heat-transfer coefficient between the vapor and the preheated brine is lower in the MSF process because the heated brine does not boil. In the ME process it does boil, and it is well known that boiling heat-transfer coefficients are significantly higher than those where the heating does not result in boiling. In using direct transfer of latent heat of condensation to latent heat of evaporation, instead of sensible heat reduction to latent heat of evaporation as in MSF, the ME process requires a much smaller brine flow than the MSF. Limiting brine concentration in the last effect to about three times that of the entering seawater, for example, requires a brine flow of only about 1.5 times that of the distillate produced. At the same time, a pump (although much smaller than the two pumps needed in MSF) is needed for each effect.

The performance ratio of ME plants is just slightly lower than the number of effects, which is determined as an optimized compromise between energy efficiency and capital cost. Six effects are typical, although plants with as many as 18 effects have been built. Further detail about ME desalination can be found in (56,59).

Vapor Compression Distillation (VC). The vapor pressure of saline water is lower than that of pure water at the same temperature, the pressure difference being proportional to the boiling point elevation of the saline water. Desalination is attained by evaporating the saline water and condensing the vapor on top of the pure water. Therefore, the pressure of the saline water vapor must be raised by the magnitude of that pressure difference, plus some additional amount to compensate for various losses. This is the principle governing the vapor compression desalination method. Moreover, as shown in the VC plant flow diagram in Figure 7, the heat of condensation of the compressed vapor is recovered internally by using it to evaporate the saline water. Additional heat recovery is obtained by transferring heat from the concentrated brine effluent and the produced fresh water (which need to be cooled down to as close to ambient conditions as possible anyway) to the feed saline water, which is thus preheated. The schematic flow diagram in Figure 7 shows a design in which the preheated seawater is sprayed onto a bank of horizontal tubes carrying condensing compressed vapor at a temperature higher than that of the seawater. The spray thus evaporates on contact with the exterior of the tube and provides the cooling needed for the internal condensation. Considering the fact that the energy required for vapor compression over a typical overall temperature difference of 4°C and a vapor compressor efficiency of 0.8 is 34 kJ / kg (easily calculated from an enthalpy balance), and that the latent heat of condensation is about 2400 kJ / kg, one can see that a small amount of compression energy enables a large amount of heat to be used internally for desalination. One can thus envisage the VC plant as a large flywheel, wheeling a large amount of energy around at the expense of a small amount needed for sustaining its motion.

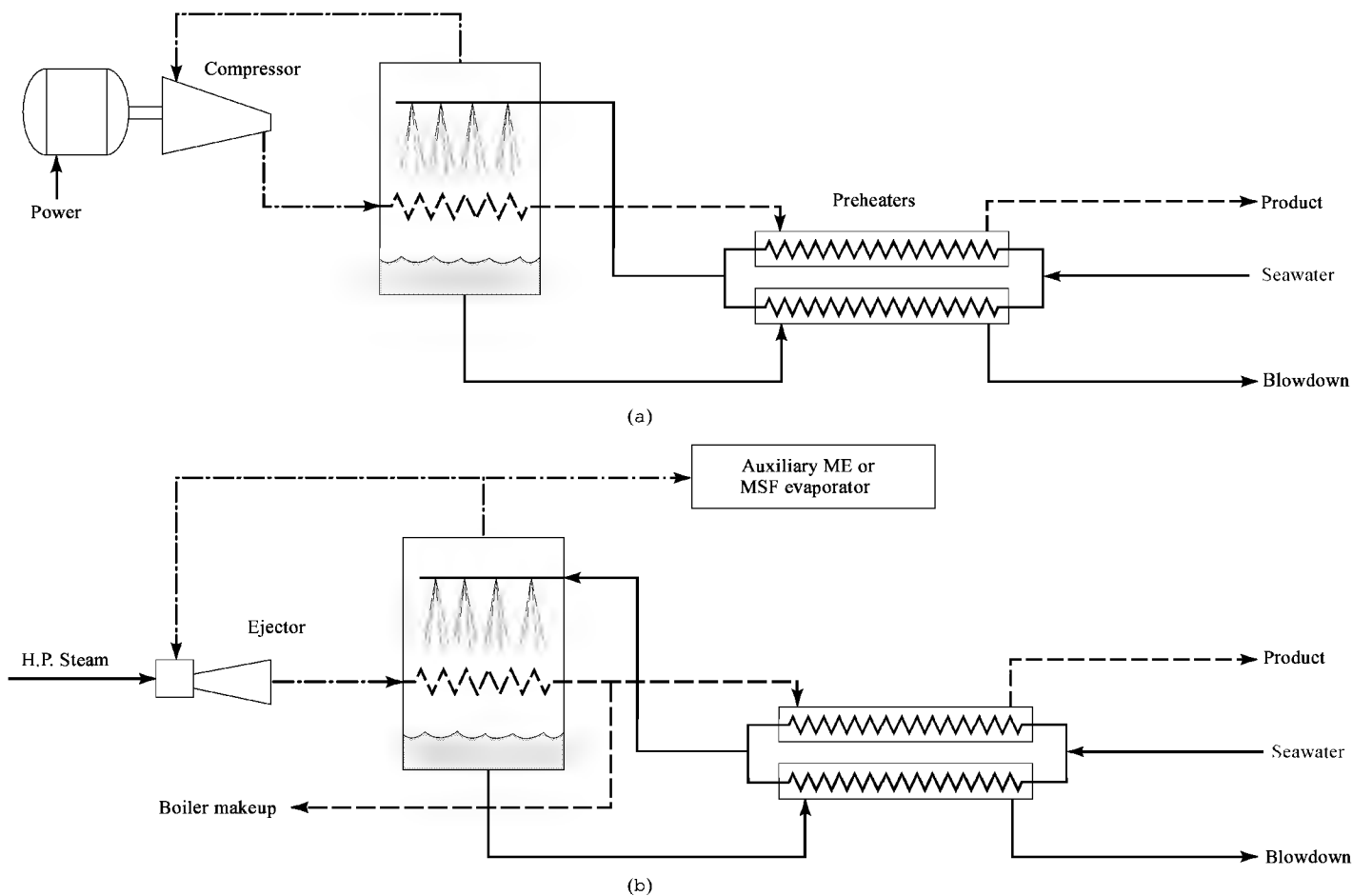


Fig. 7. Schematic flow diagram of a basic horizontal-tube vapor compression (VC) desalination plant, shown (a) with a mechanical, motor-driven compressor and (b) with a thermocompressor, using an ejector, where (—) represents vapor, (—), brine; and (—), product.

The compressor can be driven by electric motors, gas or steam turbines, or internal combustion (usually diesel) engines. The compressor can also be a steam-driven ejector (Fig. 7b), which improves plant reliability because of its simplicity and absence of moving parts, but also reduces its efficiency because an ejector is less efficient than a mechanical compressor. In all of the thermally driven devices, turbines, engines, and the ejector mentioned herein, the exhaust heat can be used for process efficiency improvement, or for desalination by an additional distillation plant. Figure 8 shows a flow diagram of the vertical-tube vapor compression process.

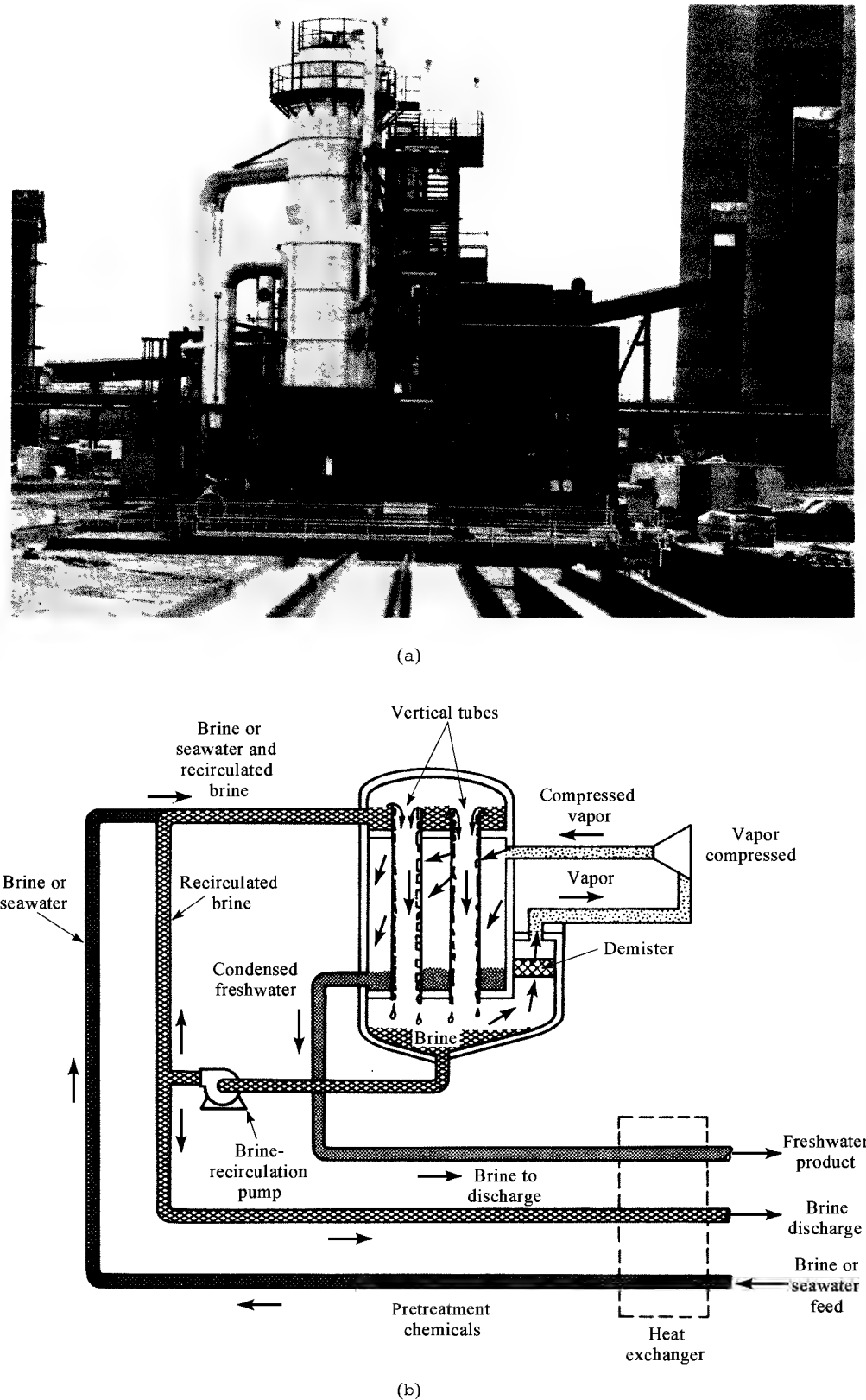


Fig. 8. Pictorial view (a) and flow diagram (b) for a vertical-tube vapor-compression process. Courtesy of Resources Conservation Co.

Figure 9 shows a multi-effect VC plant. Using more than a single effect reduces the vapor volume that needs to be compressed. Moreover, the overall required heat-transfer area is also decreased because much of the single-phase heat-transfer process in the preheater of the single-effect plant is replaced by the high heat-transfer condensation–evaporation processes in the effects. Although the multi-effect feature also increases the required compression ratio, the cost of produced water is reduced overall. An operating two-stage 700-m³ d VC brine concentration plant is shown in Figure 10.

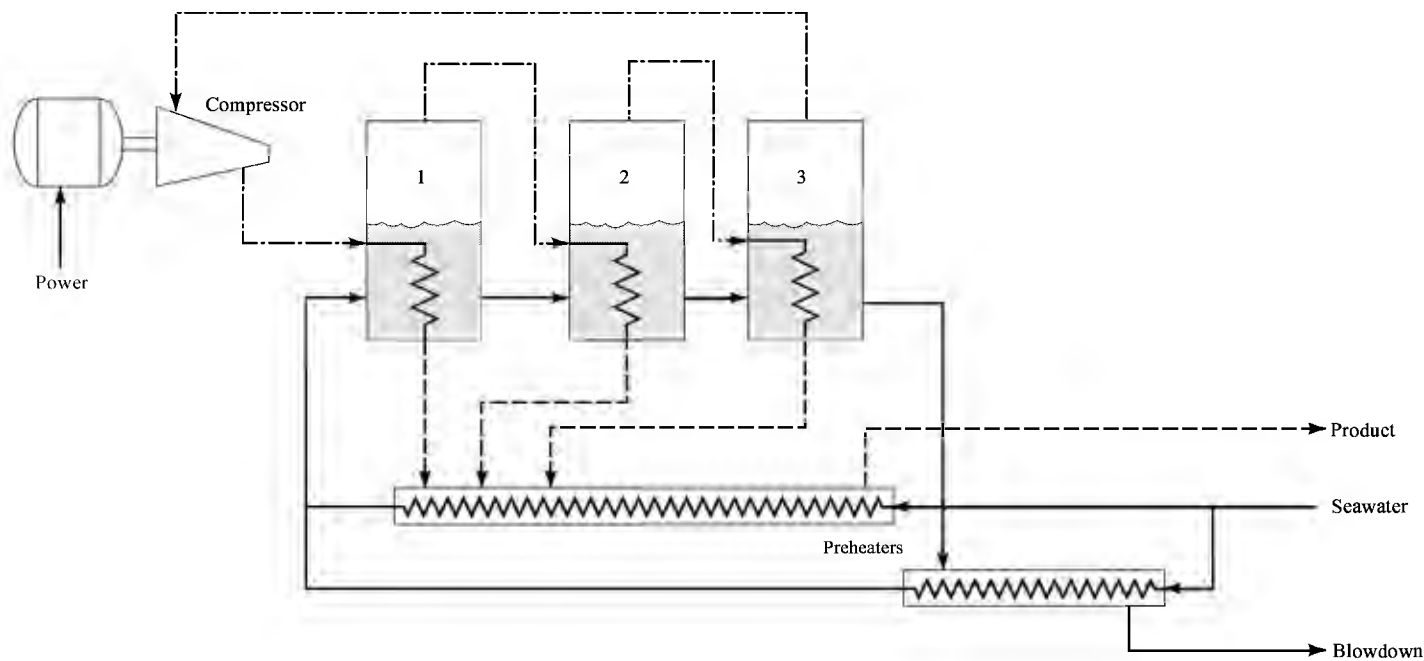


Fig. 9. Schematic flow diagram of a multi-effect vapor-compression submerged-tube desalination plant with three effects, where (---) represents vapor; (—), brine; and (—○—), product.



Fig. 10. A two-stage 700-m³/d VC brine concentration plant.

Courtesy of I.D.E. Technologies Ltd.

VC plants are of small (<4,000 m³/d) capacity and relatively high energy efficiency, the latter depending on the number of effects incorporated, eg, a two-effect plant had an energy use of 57 kJ per kg water produced. Further detail about VC desalination can be found in References 48, 55, 56, and 60.

Freeze-Desalination. It is rather well known that freezing of saline water solutions is an effective separation process in that it generates ice crystals which are essentially salt-free, surrounded by a more salt-concentrated solution. This phenomenon has significant underlying appeal for its use as a water desalination process: (1) compared with distillation, which requires sensible heating of the saline water from an ambient temperature of, eg, 25°C to about 100°C, a 75°C rise, freezing needs cooling by only about 25°C to the freezing point; (2) in distillation, a latent heat investment of about 2400 kJ/kg is needed for generating the pure water vapor, whereas the latent heat of freezing is only about 333 kJ/kg, almost 8-fold smaller; (3) operation at freezing temperatures reduces the problems of corrosion and scaling significantly; and (4) the low temperature also allows the use of less expensive construction materials.

Cooling to the freezing point can be attained by using a mechanically or thermally driven refrigeration process where a refrigerant is cooled down and used to withdraw heat from the saline water, a process known as secondary refrigerant direct or indirect freezing. Another way, shown in Figure 11, is by using the water itself as the refrigerant in the direct freezing process (61). Pretreated and deaerated seawater is precooled in a heat exchanger, where the product and blow-down brine are thus heated to ambient temperature. The cooled seawater then enters a vacuum chamber, ~500 Pa, in which the pressure is below that corresponding to saturation. The seawater therefore flashes there into vapor, just like the phenomenon in a stage of a multi-stage flash evaporation plant. All of the latent heat of evaporation is obtained from the sensible heat of the solution, thus reducing its temperature significantly and causing it to freeze into freshwater crystals. The freezing point of saline water is depressed to below that of freshwater to a degree proportional to the salt

concentration. The concentration in the crystallizer is about twice that of raw seawater, and the freezing point is then about -3.8°C . To condense the vapor, also composed of freshwater, it is compressed to a pressure that is high enough to allow its condensation on the formed ice, ie, corresponding approximately to that at the saturation temperature of pure water, 0°C .

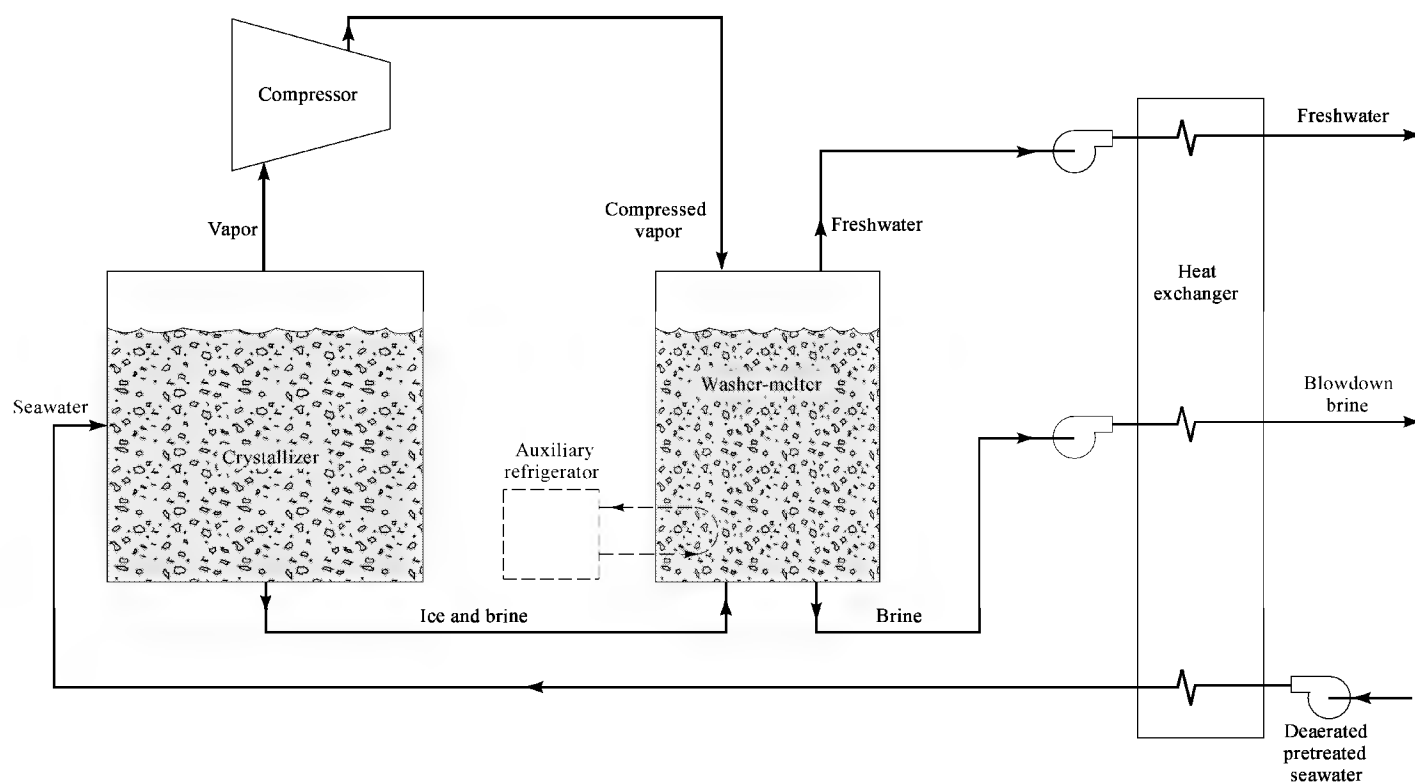


Fig. 11. Flow diagram of the vacuum-freeze (direct) vapor-compression desalination process.

The ice crystals must be separated from the saline solution surrounding them, and washed with freshwater. This is accomplished by a downward countercurrent flow of a small amount of freshwater through the ice slurry in the washer-melter unit. Keeping that unit at about 0°C limits the needed pressure rise by the compressor to only about 130–260 Pa, and an auxiliary refrigerator is often used to compensate for heat gains from the ambient and the compression.

In principle, freeze-desalination is perhaps the most energy-efficient desalination process, but it has not yet (ca 1997) reached commercial introduction for several reasons, such as the difficulty in developing efficient and economical compressors for vapor having the extremely high specific volume at the low process pressure, and difficulties in maintaining the vacuum system leak-free and in effecting reliable washing of the ice crystals. A review of freeze desalination processes is given in Reference 62.

Membrane Separation Processes

Reverse Osmosis (RO). The reverse osmosis process may be introduced by comparing it to the well-known process of filtration, which separates particulate matter from a fluid by applying pressure to the fluid and passing it through a porous membrane. Here particles larger than the pore size remain on the upstream side of the membrane and the fluid cleaned thereby flows to its downstream side. Semipermeable, very dense membranes that actually separate salt molecules (ions) from the water, by similarly keeping the salt on the upstream side and allowing the pressurized pure water to flow through the membrane, were developed in the late 1950s and early 1960s. The reverse of this process, osmosis, is well known. For example, if a membrane is placed to separate water from an aqueous salt solution, and the membrane is semipermeable (here meaning that it permits transfer of water only, not the salt components in the aqueous solution) the water will naturally migrate through this membrane into the salt solution until the thermodynamic potentials are balanced. Osmosis is, for example, the major mass transport phenomenon across living cells. The driving force for this water flux is the thermodynamic potential difference, proportional to the concentration difference between the two sides of the membrane, and is exhibited as the so-called osmotic pressure. For typical seawater at 25°C , the osmotic pressure is higher by 2.51 MPa on the freshwater side of the membrane. If a pressure higher than the osmotic pressure is applied to the concentrated solution side of the membrane, the water will move across the membrane in the reverse direction, from the saline solution side to the pure water one. This process is called reverse osmosis (and sometimes hyperfiltration), and is the basic principle underlying reverse-osmosis desalination.

Unlike filtration of particulates, the selective filtration of the water in reverse osmosis is not the result of the relationship of the membrane pore size to the relative sizes of the salt and water molecules. One way to view the process is that the very thin active surface layer of the membrane forms hydrogen bonds with water molecules and thus makes them unavailable for dissolving salt. Salt thus cannot penetrate through that layer. Water molecules approaching that pure water layer are, however, transported through it by forming such hydrogen bonds with it and in that process displacing water molecules that were previously hydrogen-bonded at these sites. The displaced water molecules then move by the forces of the hydraulic pressure difference through the pores of the remainder of the membrane, emerging at its other side. Several other mechanisms for explaining and modeling RO transport exist in the literature (63–67). It is noteworthy that RO also separates small-molecule organic solutes from the water.

The prevalent membrane configurations used in RO plants are of the spiral-wound and hollow fiber types. The basic spiral-wound-type module (Fig. 12a) is made of two sheets placed upon each other and rolled together in a spiral around a cylindrical perforated tube. One of the sheets is in the form of a sandwich typically composed of five layers bonded together along three edges. The two outer layers are the semipermeable membranes. Each of them is backed by a porous material layer for mechanical strength, and the central layer is a thicker porous material layer that collects and transports the produced fresh water. The second sheet is a porous mesh spacer through which the high-pressure saline water feed is passed in an axial direction. Product water separates from the saline solution and permeates through the two adjacent semipermeable membranes into the central product-water carrying layer, which conducts it spirally to the unbonded edge of the sandwich to the central perforated tube. The semipermeable membranes are typically made from cellulose acetate, and more recently have been made from composites of polyamide polymers on polysulfone support membranes. Fig. 12b shows a version of a contemporary commercial element of that type.

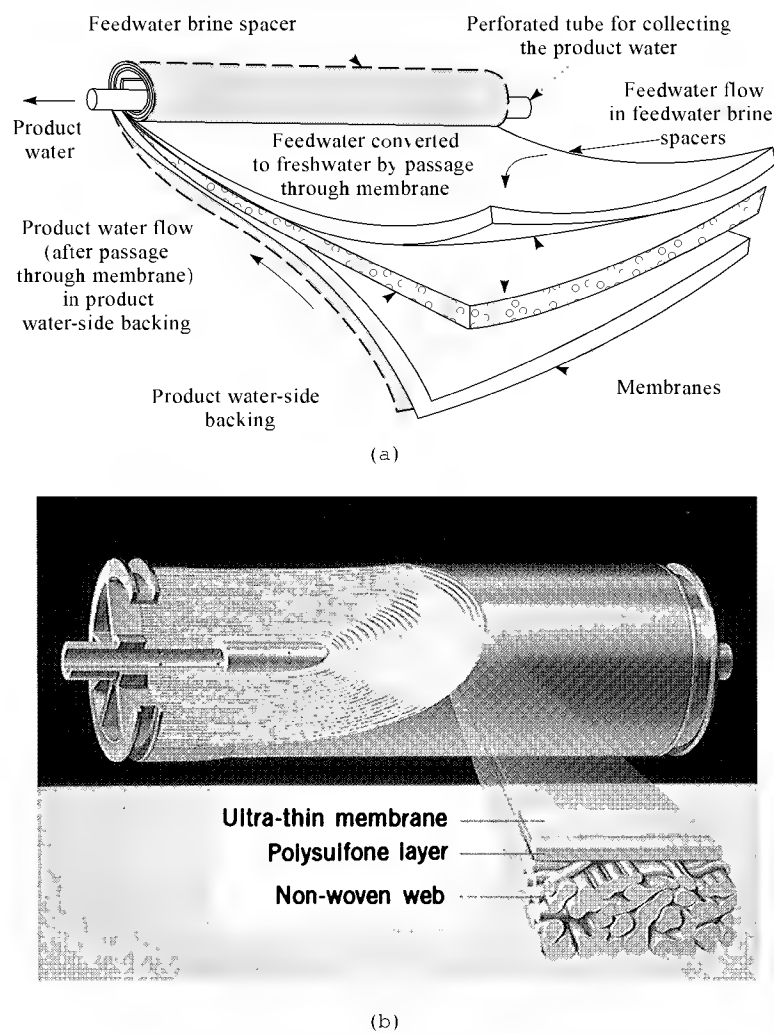


Fig. 12. A spiral-wound reverse osmosis membrane element: (a) schematic depiction; (b) cross section of a spiral-wound thin-film composite RO Filmtec membrane element (40% conversion). Courtesy of Dow Chemical Co.

Hollow-fiber modules have a configuration similar to a shell-and-tube heat exchanger, with the fibers taking the place of the tubes. A very large number of typically 25–250 μm outside-diameter semipermeable hollow fibers (wall thickness typically 5–50 μm) are bundled together and placed in a saline water-pressure vessel. The hollow core of each fiber is sealed on one end. The pressurized saline water is brought into the module through a central porous feed tube (Fig. 13) to circulate on the exterior surface of the fibers, and water permeates through the fiber wall into its hollow core, through which it flows to a permeate collection manifold at the open end of the fiber bundle. The increasingly concentrated saline water flows radially and is discharged at the exterior shell of the bundle. The hollow fibers are typically made of polyamide or cellulose triacetate, and offer about 20-fold more surface (separation) area per unit volume than the spiral-wound configuration.

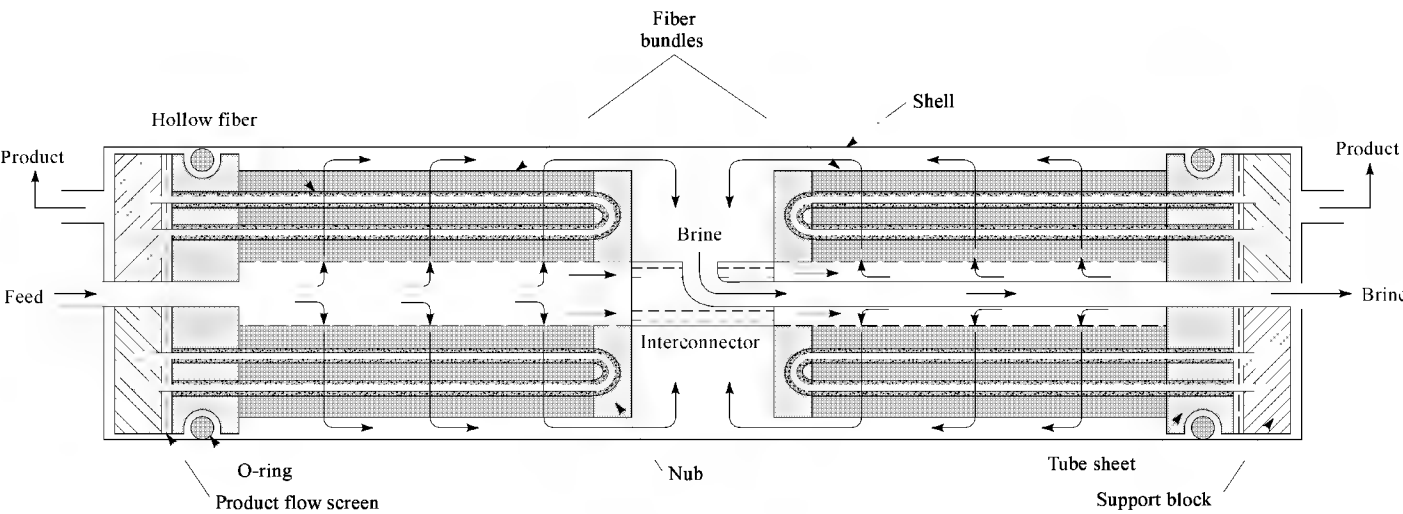


Fig. 13. A hollow-fiber reverse osmosis membrane element. Courtesy of DuPont Permasep. In this twin design, the feedwater is fed under pressure into a central distributor tube where half the water is forced out radially through the first, ie, left-hand, fiber bundle and thus desalted. The remaining portion of the feedwater flows through the interconnector to an annular feed tube of the second, ie, right-hand, fiber bundle. As in the first bundle, the pressurized feedwater is forced out radially and desalted. The product water flows through the hollow fibers, collects at each end of the element, and exits there. The concentrated brine from both bundles flows through the concentric tube in the center of the second bundle and exits the element on the right.

The basic approximate equation for the separation process gives the water flux, $\dot{m}_w''$ (kg m² s) across an RO membrane, in the absence of fouling, as

$$\dot{m}_w'' = K_{pe} K_{cf} [(P_f - P_p) - (\pi_f - \pi_p)] \tag{12}$$

where K_{pe} = the water permeability constant of the membrane (in kg m² sPa), typically increasing strongly as the temperature rises: a plant designed to operate at 20°C may produce up to 24% more water if the water temperature is 28°C; K_{cf} = the Compaction Correction Factor (dimensionless) which corrects for the fact that the flux is reduced as a result of densification of the barrier layer (a phenomenon similar to creep) of the membrane, and which increases with the operating pressure and temperature. It is often calculated by the relationship

$$K_{cf} = BC(T)C(P)C(t) \tag{13}$$

where B is a constant, $C(T)$ represents the temperature dependence of the compaction correction factor for the particular membrane of interest; $C(P)$ represents its pressure dependence: whereas a higher pressure difference across the membrane is shown in equation 12 to increase the water flux, higher feed pressure, (P_f) , also tends to compact the membrane and thus reduce its water flux, typically according to

$$C(P) = P_f^n \tag{14}$$

where n is a negative number, and where the time dependence $C(t)$ is represented by

$$C_{cf} \equiv t^m \tag{15}$$

where t is the operating time (eg, in days) and m is a negative number depending on the membrane; P = water or saline solution pressure (in Pa); π = osmotic pressure (in Pa); and the subscripts f and p pertain to the saline feed water and to the desalted product water, respectively.

The required membrane area A can be estimated by

$$A = \frac{\dot{m}_p}{\dot{m}_p'' f} \tag{16}$$

where $\dot{m}_p$ is the freshwater mass-production rate of the plant, and f ($0 < f < 1.0$) is the area utilization factor that corrects for the fact that the membrane surface is incompletely in contact with the saline water feed stream owing to the porous mesh and other devices, such as turbulence promoters, placed in the feed stream path; in a good design $f < 0.9$.

Examination of equation 12 shows that water separation rate increases with the Water Permeability Constant K_{pe} . Unfortunately, the salt flux across the membrane also does, resulting in a more salty product. An approximation for this salt flow is

$$\dot{m}_s = K K_s (C_{fm} - C_p) \tag{17}$$

where $\dot{m}_s$ = salt mass transfer rate across the membrane, kg s⁻¹; K = a proportionality constant, dimensionless; K_s = the salt permeation constant, kg s⁻¹, which increases with pressure and temperature.

The salinity of the product water, C_p , can be estimated by the formula

$$C_p = K_{cp}(1 - \eta) \overline{C} \tag{18}$$

where K_{cp} = the concentration polarization coefficient, $= C_{fm} / \overline{C}$, a measure of the increase of the feedwater salinity at the membrane wall beyond that of the bulk solution; C_{fm} = the salt concentration at the membrane wall; $\overline{C}$ = the bulk salinity of the saline water feed, $\approx (C_f + C_r) / 2$, where C_r = the salt concentration of the reject brine; and η = the salt rejection factor, $(\text{amount of salts rejected by the membrane}) / (\text{amount of salts in the brine feed})$.

The pressure to be used for reverse osmosis depends on the salinity of the feedwater, the type of membrane, and the desired product purity. It ranges from about 1.5 MPa for low feed concentrations or high flux membranes, through 2.5–4 MPa for brackish waters, and to 6–8.4 MPa for seawater desalination. In desalination of brackish or sea water, typical product water fluxes through spiral-wound membranes are about 600–800 kg m⁻² d at a recovery ratio RR of 15% and an average salt rejection of 99.5%, where

$$RR = \frac{\dot{m}_p}{\dot{m}_f} \simeq 1 - \frac{C_f}{C_r} \tag{19}$$

The fluxes in hollow-fiber membranes used in seawater desalination are 20–30-fold smaller, but the overall RO system size does not increase because the hollow-fiber membranes have a much larger surface area per module unit volume. In use with seawater, their RR is about 12–17.5% and the salt rejection ratio is up to 99.5%.

Because the concentrated reject brine is still at high pressure, it is possible to recover energy by passing this brine through hydraulic turbines and thus reduce the overall energy consumption by up to 20%. The energy requirements of seawater RO desalination plants with energy recovery are about 5–9 kWh, or 18–33 MJ, of mechanical or electric power per m³ fresh water produced. In comparison, the MSF desalination process requires about 120–280 MJ of heat and about 15 MJ of mechanical / electric power (for pumping and auxiliaries) per m³. The energy requirement of the RO process is thus smaller than that of the MSF process even if the RO energy requirement is multiplied by the thermal-to-mechanical (or electrical) power conversion factor of 3–4. The specific exergy (or availability, the thermodynamic potential of useful work) consumption of the MSF process using 120°C steam is about 2–3-fold higher than that of the RO process, but becomes comparable in magnitude if the steam temperature is lowered to 80°C.

The life of membranes is affected by gradual chemical decomposition or change. For example, cellulose acetate membranes hydrolyze with time. The rate of hydrolysis has a steep minimum at a solution pH of 4.5–5.0, and increases drastically with temperature.

Membranes are very susceptible to fouling by solids (soil, sand, colloids, precipitated salts and corrosion products, and microbial growth) and to deterioration in their selectivity caused by various species present in the saline water. Very careful pretreatment of the feedwater is therefore necessary. It typically consists of clarification, filtration, chlorination for destroying organic matter and microorganisms, removal of excess chlorine to prevent membrane oxidation, and dosing with additives to prevent calcium sulfate scaling and foam formation. Periodical chemical or mechanical cleaning is also necessary. Pretreatment and cleaning are significant and increasing fractions of the RO plant capital and operating costs.

The plant scheme shown in Figure 14a is generally used in brackish-water conversion; that in Figure 14b is utilized in seawater-conversion plants.

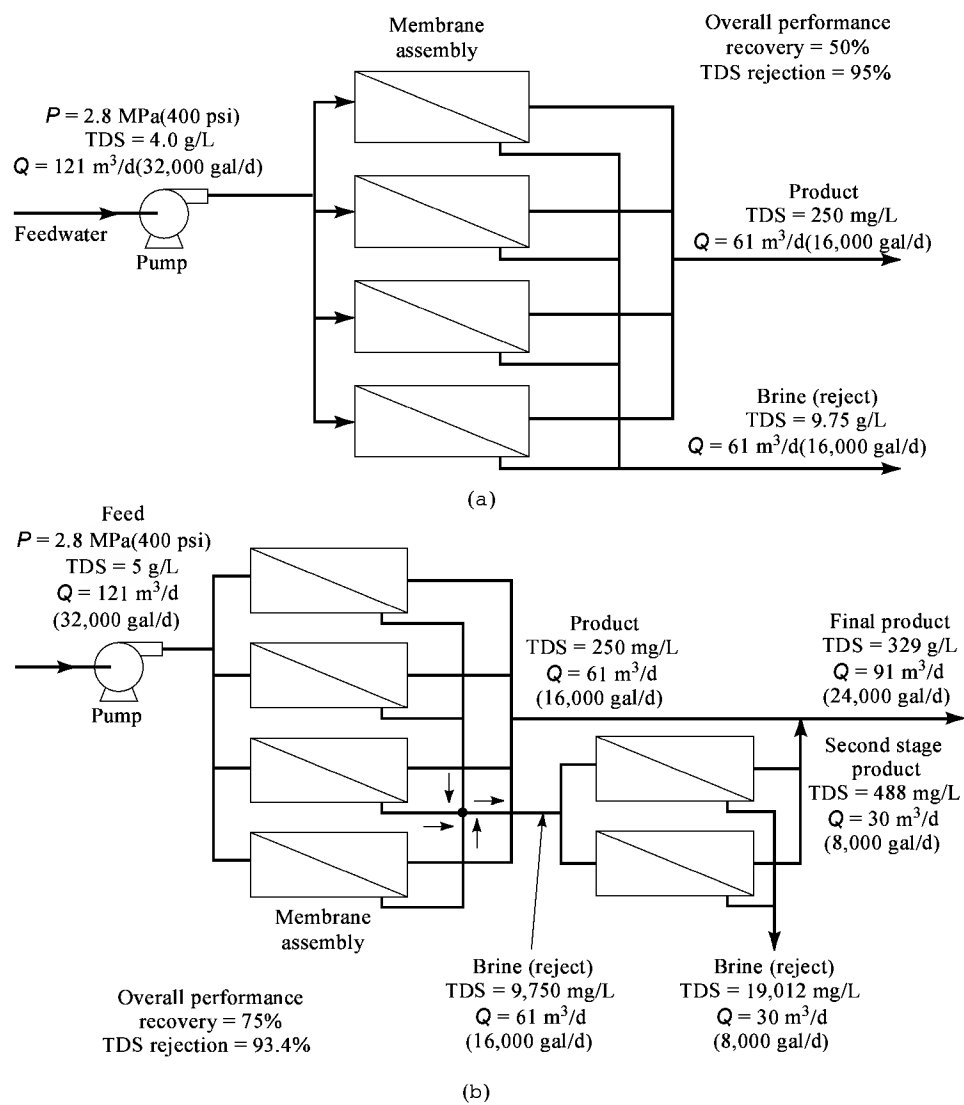


Fig. 14. Reverse osmosis plant configurations. (a) A single-stage RO plant configuration with parallel membrane assemblies, with some typical operational parameters given. (b) A multistage RO plant configuration with reject staging. This configuration is referred to as having a cascade, tapered, or pyramid configuration. Two stages are shown here; three-stage units, with recoveries of 85–90%, are also used.

Further detail about RO desalination can be found in References 60–64.

Electrodialysis. In electrodialysis (ED), the saline solution is placed between two membranes, one permeable to cations only and the other to anions only. A direct electrical current is passed across this system by means of two electrodes, causing the cations in the saline solution to move toward the cathode, and the anions to the anode. As shown in Figure 15, the anions can only leave one compartment in their travel to the anode, because a membrane separating them from the anode is permeable to them. Cations are both excluded from one compartment and concentrated in the compartment toward the cathode. This reduces the salt concentration in some compartments, and increases it in others. Tens to hundreds of such compartments are stacked together in practical ED plants, leading to the creation of alternating compartments of fresh and salt-concentrated water. ED is a continuous-flow process, where saline feed is continuously fed into all compartments and the product water and concentrated brine flow out of alternate compartments. The flow along the membranes also improves the mass transport there, and the separators between the membranes are constructed to provide good flow distribution and mixing on the membrane surfaces. Membrane sizes are often about $0.5 \times 1 \text{ m}$, spaced about 1 mm apart. Many types of polymers are used to manufacture these ion-exchange-selective membranes, which are often reinforced by strong fabrics made of other polymers or glass fibers.

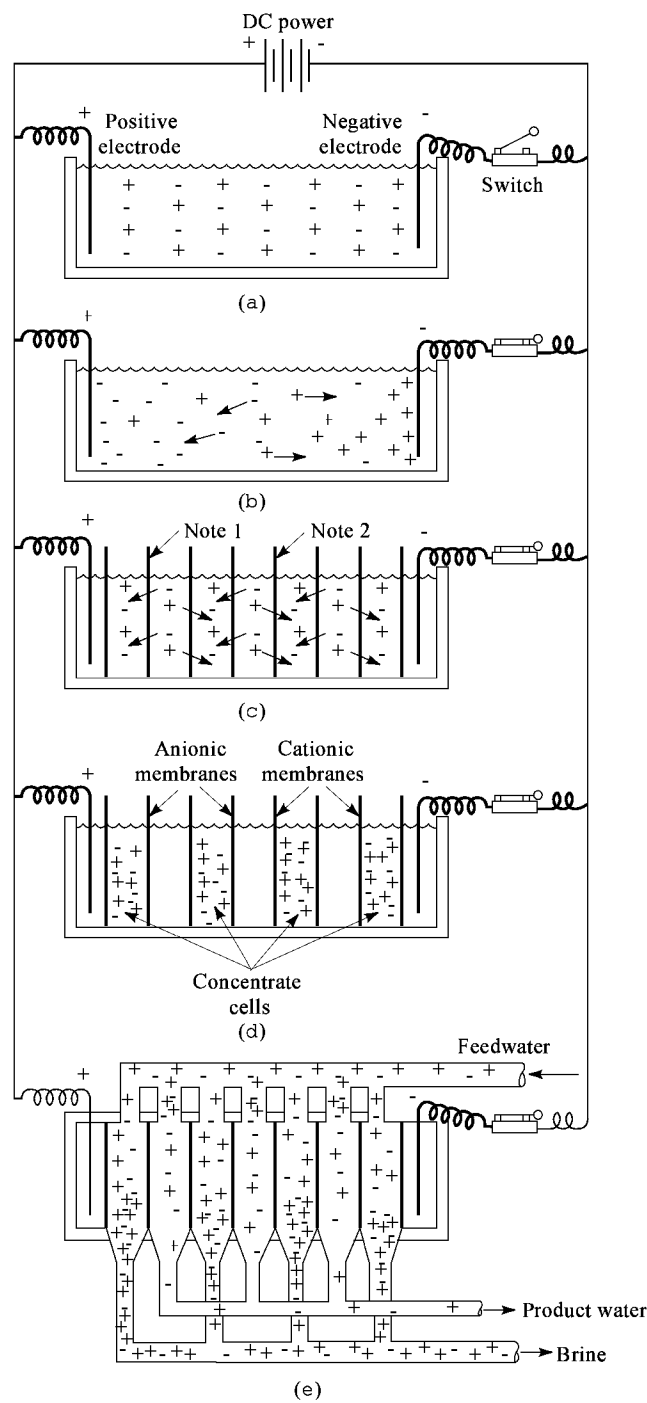


Fig. 15. Ion movements in the electrodialysis process. Courtesy U.S. Agency for International Development. (a) Many of the substances which make up the total dissolved solids in brackish water are strong electrolytes. When dissolved in water, they ionize; ie, the compounds dissociate into ions which carry an electric charge. Typical of the ions in brackish water are Cl^- , Na^+ , HCO_3^- , Mg^{2+} , SO_4^{2-} , and Ca^{2+} . These ions tend to attract the dipolar water molecules and to be diffused, in time, fairly evenly throughout a solution. (b) If two electrodes are placed in a solution of ions, and energized by a battery or other direct-current source, the current is carried through the solution by the charged particles and the ions tend to migrate to the electrode of the opposite charge. (c) If, alternatively, fixed-charge membranes, which are selectively permeable to ions of the opposite charge, are placed in the path of the migrating ions, the ions are trapped between the alternate cells formed. A positively fixed-charge (anionic) membrane allows negative ions to pass but repels positive ions. A negatively fixed-charge (cationic) membrane allows positive ions to pass, but repels negative ions. (d) If this continues, almost all the ions become trapped in the alternate cells, which lack ions, have a lower level of dissolved constituents, and have a high resistance to current flow. (e) The phenomenon illustrated above is used in electrodialysis to remove ions from incoming saline water on a continuous basis. Feedwater enters both the concentrate and product cells. Up to about half of the ions in the product cells migrate and are trapped in the concentrate cells. Two streams emerge from the device: one of concentrated brine and the other with a much lower concentration of dissolved solids (product water).

Since membrane fouling could quickly render the system inefficient, very careful and thorough feedwater pretreatment similar to that described in the section on RO, is required. Some pretreatment needs, and operational problems of scaling are diminished in the electrodialysis reversal (EDR) process, in which the electric current flow direction is periodically (eg, 3–4 times h) reversed, with simultaneous switching of the water-flow connections. This also reverses the salt concentration buildup at the membrane and electrode surfaces, and prevents concentrations that cause the precipitation of salts and scale deposition. A schematic and photograph of a typical ED plant are shown in Figure 16.

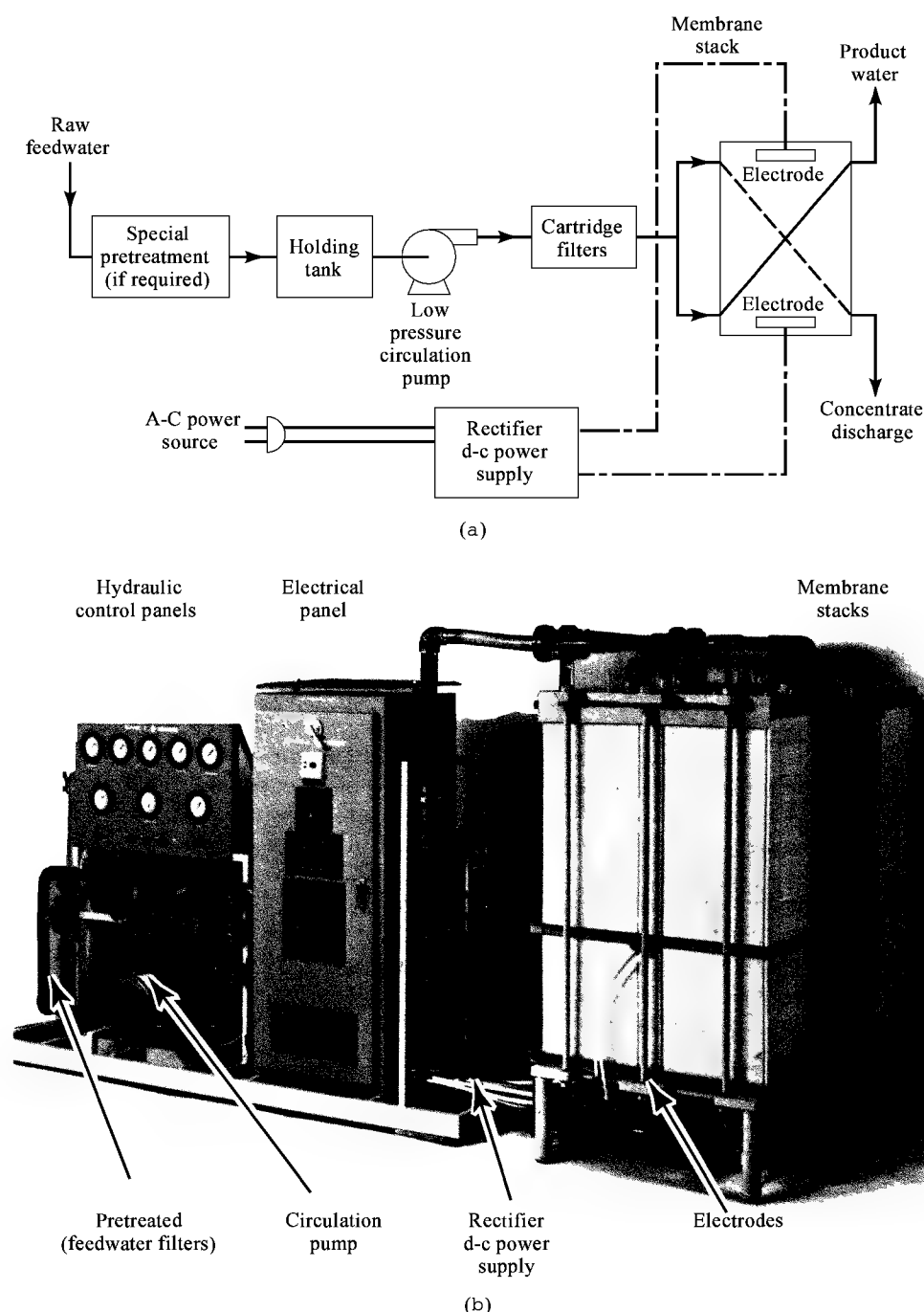


Fig. 16. Schematic (a) and basic components (b) of an electrodialysis unit.

Courtesy of Ionics.

The voltage used for electrodialysis is about 1 V per membrane pair, and the current flux is of the order of 100 A m^2 of membrane surface. The total power requirement increases with the feedwater salt concentration, amounting to about 10 MW per m^3 product water per 1000 ppm reduction in salinity. About half this power is required for separation and half for pumping. Many plant flow arrangements exist, and their description can be found, along with other details about the process, in References 68 and 69. Many ED plants, as large as $15,000 \text{ m}^3 \text{ d}$, are in operation, reducing brackish water concentration typically by a factor of 3–4.

Solar Desalination

The benefits of using the nonpolluting and inexhaustible energy of the sun for water desalination are obvious. Furthermore, many water-poor regions also have a relatively high solar flux for a large fraction of the time. The major impediment to the use of solar energy is economical: the diffuse nature of solar energy, even at its highest, dictates the need for constructing a large solar energy collection area. For example, typical annually averaged solar energy inputs in sunny regions range from $17\text{--}23 \text{ MJ m}^2 \text{ d}$. The latent heat of evaporation of water is about 2.407 MJ kg at 40°C , and assuming a single-effect solar still efficiency of 50% (which is the upper practical limit for conventional designs), the still would produce at most about $3.5\text{--}4.8 \text{ (kg fresh water) m}^2 \text{ d}$, or a $208\text{--}286\text{-m}^2$ solar still would be required to produce 1 m^3 of fresh water per day. More realistic still efficiencies reduce the production rate to about $2 \text{ (kg fresh water) m}^2 \text{ d}$ and increase the still area requirement for producing 1 m^3 of fresh water per day to about 500 m^2 . Consequently, whereas the energy input for solar desalination is free, the capital investment may be high and thus the cost of solar-desalted water may be higher than that of water desalted by processes which indeed require nonfree energy input (such as fuel) but have lower capital costs. Solar desalination technology is mature and reliable, and thus becomes competitive with other desalination processes in regions or circumstances where the solar energy input and cost of fuel or of environmental penalties of its use are high, and the construction costs of the solar plant are low.

A typical solar still, shown in Figure 17, consists of a saline water container in which the water is exposed to the sun and heated by it. The temperature rise to above ambient causes net evaporation of the saline water, thus separating pure water vapor from the solution. The transparent cover of the still serves several important functions: it prevents vapors from escaping the still; exposed to air at temperatures lower than the heated saline water in the basin, it serves as a condenser for the vapor; given proper inclination and geometric configuration, it channels the condensed fresh water to product collection troughs; it reduces heat loss from the warm saline water to the outside; and it prevents dirt from entering the still. However, it also must be as transparent as possible to solar radiation so that maximal heat gain by the still can be accomplished.

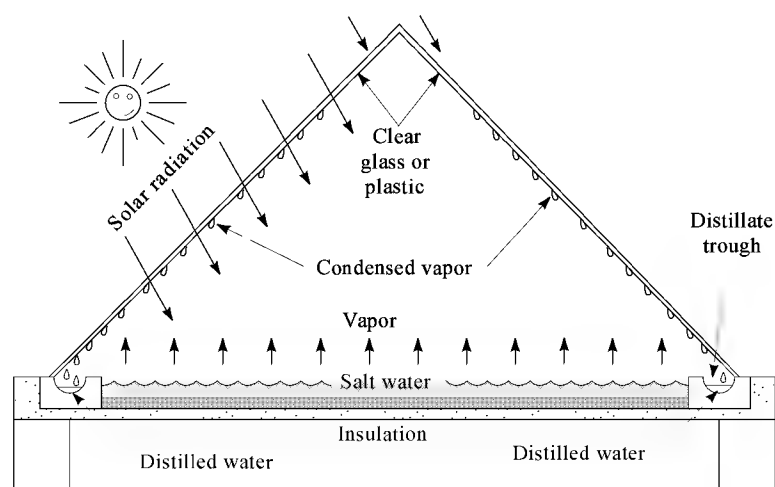


Fig. 17. A typical basin-type solar still.

Because water of depths below about 2 m does not absorb much solar radiation directly, the radiation is absorbed and converted to heat primarily in the basin floor, which thus should have high radiative absorptance in the solar radiation spectrum. It is also noteworthy that if the still is designed to have low heat losses to the ambient, and if the ambient temperature drops, distillation will continue for some time even in the absence of solar energy input, because the saline water may remain warmer than the condensing glass surface and thus continue evaporating.

Solar stills of the type depicted in Figure 17, in many sizes and constructional variants, have been built and used successfully in many countries in the world. They are simple, easy to construct, reliable, and require very little maintenance.

Because the heat of condensation in single-effect stills of the type shown in Figure 17 is lost to the ambient, it is obvious that more energy-efficient operation would be achieved if a multi-effect design could be implemented, in which the heat of condensation is used to evaporate additional saline water. There are several problems with the design implementation of this concept. The principal one is the rather low available temperature driving force for separation in even a single-effect still: the condensation temperature is the ambient, say at 25°C, whereas the diurnal average temperature of the saline water in the still is around 40°C, only 15°C higher. Another difficulty is configurational, owing to the need for adding effects without at the same time obstructing solar radiation incidence on the saline water. A number of such stills were built and tested successfully, but are not yet (ca 1997) commercially competitive.

Solar stills integrate the desalination and solar energy collection processes. Another approach to solar desalination is to use separately a conventional desalination process and a solar energy supply system suitable for it. Any compatible desalination and solar energy collection processes could be used. Distillation, such as MSF or ME, can be used with heat input from solar collectors or concentrators (70) or from solar ponds (71–73). Net average solar energy conversion efficiencies of solar collectors (74,75) are about 25%, and of solar ponds (71–73,76,77), 18%, similar to that of solar stills, but the MSF or ME plants can operate at performance ratios of 10 or more, thus basically increasing the freshwater production rate by at least tenfold, or reducing the required solar collection area by at least tenfold for the same production rate.

Still in the R&D stage, open-cycle ocean–thermal energy conversion (OTEC) plants using surface condensers produce desalted water (78,79). The concept is illustrated in Figure 18. The warmer surface ocean water is pumped into a flash evaporation vessel from which the air has been evacuated. The vapor generated by flashing passes through a turbine and thus produces power, and is then condensed in a deaerated surface condenser cooled by colder (typically by about 20°C) seawater pumped to it from the ocean depths. The condensed vapor is freshwater, which can then be pumped out and used. In addition to requiring no fuel or other form of nonrenewable energy, OTEC and solar pond desalination have an advantage in that thermal energy storage is naturally included in the large thermal mass of the ocean or pond, while such storage may need to be constructed where the heat to the desalination process is supplied by solar collectors. It is noteworthy that fresh water can be produced by the OTEC-type cycle even without power production.

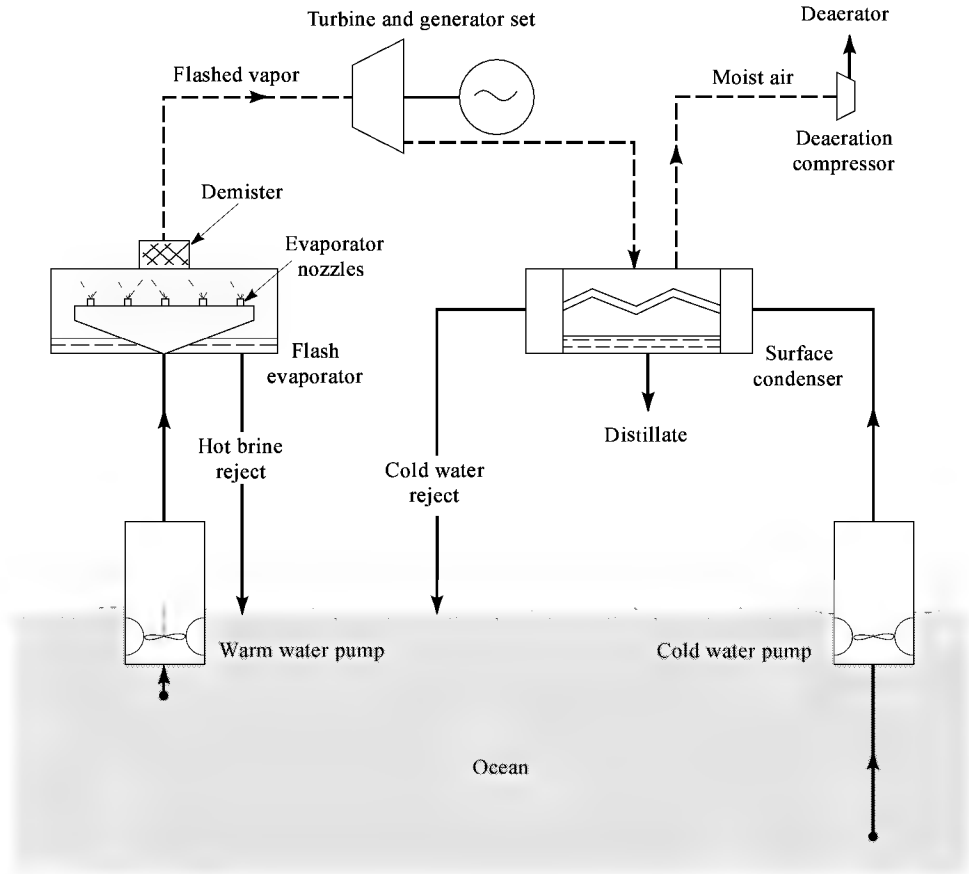


Fig. 18. Flow schematic of an open-cycle ocean-thermal energy conversion (OTEC) and desalination plant (77), where ($\longrightarrow$) represents liquids; ($\rightarrow$), vapors or gases.

Solar or wind energy can also be used for desalination processes which are driven by mechanical or electrical power, such as VC, RO, and ED. The solar heat can be used to generate the required power by a variety of means. Figure 19 shows the flow diagram of a reverse osmosis plant in which the pumps are driven by photovoltaic solar cells, built by the French Atomic Energy Commission at Cadarache, France (80). A similar system, powered by an 8-kW array of photovoltaic modules has also been installed in Jeddah, Saudi Arabia (81). A wind-powered vapor compression desalination demonstration plant using a 60-kW (max) electric power output wind turbine and producing up to 48 m³ d of freshwater was installed on the island of Borkum, and a prototype plant using a 300 kW (max) electric power wind turbine and producing up to 360 m³ d of freshwater was installed on the island of Rügen, both in the North Sea of Germany (82). The electricity produced by the wind-turbine-generator system is used to drive the VC plant compressor and to provide resistance heat to the VC system. Variable water output by the plant resulting from variations in wind velocity is accommodated easily by product water storage. The reported initial results have shown efficient and reliable operation.

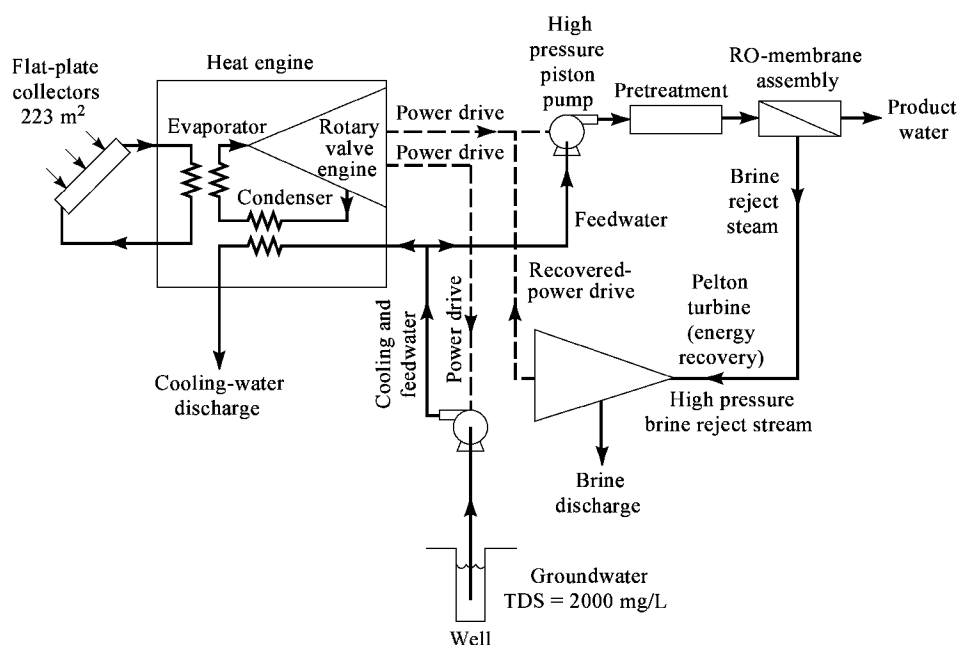


Fig. 19. Flow diagram of the experimental solar RO unit at Cadarache, France. Feedwater flow = 1.38 L s, product water flow = 0.69 L s, energy consumption without recovery = 0.89 kW · h m³ (76).

Courtesy of Commissariat d'Energie Atomique, France.

Summary and Future Prospects

Rising population, standards of living, and water pollution are gradually diminishing the amounts of naturally available freshwater of good quality, whereas the demand is relentlessly increasing. Manufactured water in the form of water desalination is today making a considerable contribution to the world's potable water supply as well as water for industry, ranging from boiler feedwater for the power industry to ultrapure water for the electronics industry. The process industries are also looking at desalination technology either for better process water, or as means of conservation through recycling of wastewaters. In some industries, it is considered a technique for recovering some process chemicals from water and lowering the cost of purification of polluting discharges. Some desalinated water is also used in agriculture, but only to a limited extent because of cost. The technology is improving both in cost-performance and reliability, as evidenced not only by the rapid growth in worldwide desalination capacity, but also by the encouraging observation that costs of desalted water have remained the same despite inflation.

Although desalination technologies are diverse, MSF has been for some time, and will remain well into the next century, the main process for desalination of seawater. Inroads are being made by the multi-effect processes and, in particular, by the low temperature ME processes.

Important advances in the membrane field were responsible for the commercialization of seawater RO. This technique both saves energy and offers much-reduced plant construction time. Compared with distillation, RO treatment of seawater has had a meteoric success beginning in the early 1990s, and further growth is expected. Considering brackish-water RO and EDR as well, the membrane processes have become an integral and growing part of desalination. In fact, their combined growth appears to exceed that of distillation.

The steady growth in water consumption around the world without any real growth of available water resources is focusing increasing attention on water reuse. Desalination technology is expected to be an integral part of all water reuse schemes. Water reuse is already receiving serious consideration ca 1997, and within the next ten years there will be water-reuse schemes in virtually all areas of water use, ie, not only in water-scarce regions, but also in water-abundant regions such as the industrial northeast of the United States. Municipal authorities will become intimately involved in conversion of their wastewater in the future.

Much progress has been effected in desalination technology, but all desalination processes remain energy-intensive. Although energy costs have prompted the consideration of renewable energy resources as sources for desalination, these sources are unlikely to become significant, except in certain limited circumstances, until there are significant unforeseen breakthroughs in either energy conversion or desalination technologies.

Despite the cost of desalination technology, it has made water available in places where it was not before. Not only has water become available in these places, but the quantities available have also opened prospects for industrial development. This has led to important improvements in the standard of living with prospects for even further improvements in countries fortunate enough to be able to meet the cost of the technology.

BIBLIOGRAPHY

"Water, Demineralization" in *ECT* 1st ed., Suppl. 1, pp. 908–930, by E. R. Gilliland, Massachusetts Institute of Technology; "Water Supply and Desalination" in *ECT* 2nd ed., Vol. 22, pp. 1–65, by D. F. Othmer, Polytechnic Institute of Brooklyn; in *ECT* 3rd ed., Vol. 24, pp. 327–366, by R. Bakish, Bakish Materials Corp.

1. A. LaBastille and T. Spiegel, *Nat. Geograph.*, 652 (Nov. 1981).

2. National Acid Precipitation Assessment Program (NAPAP), *1992 Report to Congress*, NAPAP, Washington, D.C., 1992.
3. *Bus. Week* (Feb. 22, 1969).
4. R. Hurwitz, personal communication, New York City Dept. of Environmental Protection, Bureau of Water Supply, 1996; A. O. Sulzberger, *N.Y. Times* (Aug. 10, 1981).
5. J. Opie, *Ogallala Water for a Dry Land*, University of Nebraska Press, Lincoln, Neb., 1993.
6. California Department of Water Resources, *California Water Plan Update*, Bulletin 160-93, Oct. 1994; W. S. Swain, USGS, personal communication, U.S. Geological Survey, Sacramento, CA, 1996.
7. J. Koenig, *J. Am. Water Works Assoc.* **51**, 845 (1959).
8. M. D. Farrell, *Int. Desalin. Water Reuse Q.* **6**(1), 20–29 (1996).
9. J. Dahme, personal communication, Philadelphia Water Dept., 1996.
10. W. B. Solley, R. R. Pierce, and H. A. Pedman, *Estimated Use of Water in the United States in 1990*, U.S. Geological Survey Circular 1081, USGS, Denver, Colo., 1993.
11. U.S. Dept. of Agriculture Forest Service, *RPA Assessment of the Forest and Rangeland Situation in the U.S., 1983*, and *1993 Update*, Forest Resources Reports No. 26 and 27, Washington, D.C.
12. D. G. Argo, *Proceedings Second Annual Conference*, National Water Supply Improvement Association, 1974.
13. W. R. Mills, Jr., *Orange County Water District Wastewater Reclamation, Talbert Barrier and Recharge Project*, Annual Report 91-121 for the California Regional Water Quality Control Board, Santa Ana Region, 1995.
14. U.S. Environmental Protection Agency (EPA), *National Quality Inventory*, 1994 Report to Congress, Report EPA841-R-95-05, Washington, D.C., 1995.
15. T. Gilchrist and R. I. Saunders, *Int. Desalin. Water Reuse Q.* **5**(4), 50–54, 1995.
16. A. Bruno-Vega and K. S. Thomas, *Int. Desalin. Water Reuse Q.* **5**(2), 25–31, 1995.
17. K. Wangnick, *Proceedings of an IDA World Congress on Desalination and Water Science*, 2, Abu Dhabi, International Desalination Association, Topsfield, Mass., 1995, pp. 17–26.
18. *Water Desalination in Developing Countries*, U.N. Publication 64 11 B 5-1964, United Nations New York, 1964.
19. R. L. Meeks, Federal R & D Funding by Budget Function, Fiscal years 1995–1997, Report NSF 97–301, National Science Foundation, Arlington, Virg., 1997.
20. S. F. Singer, *Environ. Sci. Technol.* **3**, 197 (1969).
21. *Chem. Eng.*, 23 (May 3, 1982).
22. W. M. Wagner and D. R. Finnegan, *Chem. Eng.*, 71 (Feb. 7, 1983).
23. F. B. Leitz and E. I. Ewoldsen, *Desalination* **24**, 321 (1978); F. B. Leitz, personal communication, U.S. Bureau of Reclamation, Denver, Colo., 1996.
24. Ernst & Young LLP, *1994 National Water and Wastewater Rate Survey*, Washington, D. C., 1994.
25. C. A. Pappas, *Int. Desalin. Water Reuse Q.* **3**(4), 34–39 (1993).
26. *The United Nations Organizations and Water*, Briefing note, United Nations, New York, 1993.
27. *Report of the United Nations Water Conference, Mar del Platta*, Publ. No. 77.II.A.12 United Nations, 1977.
28. United Nations Committee on Natural Resources, *Report of the Secretary General on Strategies and Measures for the Implementation of the Mar del Platta Action Plan in the 1990s*, 1991.
29. United Nations, *Proceedings of an International Symposium on Water Resources in the Middle East*, University of Illinois at Urbana-Champaign, 1993.
30. *Proceedings, First Desalination Congress of the American Continent*, Mexico, Elsevier Publishing Co., Amsterdam, the Netherlands, 1976; R. Bakish, ed., *Proceedings, International Congress on Desalination and Water Reuse, Tokyo*, International Desalination and Environmental Association, Teaneck, N.J., 1977; R. Bakish, ed., *Proceedings, International Congress on Desalination and Water Reuse, Nice*, International Desalination and Environmental Association, Teaneck, N.J., 1979, *Proceedings, International Congress on Desalination and Water Reuse Bahrain*, International Desalination and Environmental Association, Teaneck, N.J., 1981; *Proceedings, First World Congress on Desalination and Water Reuse, Florence*, International Desalination and Environmental Association (with EFChE), Teaneck, N.J., 1983; *Proceedings, Second World Congress on Desalination and Water Reuse, Bermuda*, International Desalination Association, Topsfield, Mass., 1985; *Proceedings, Third World Congress on Desalination and Water Reuse, Cannes*, International Desalination Association (with Soc. de Chimie Industrielle and Assoc. Italiana d'Ingegneria Chimica), Topsfield, Mass., 1987; *Proceedings, Fourth World Congress on Desalination and Water Reuse, Kuwait*, International Desalination and Environmental Association, Topsfield, Mass., 1989; *Proceedings, IDA and WRPC World Conference on Desalination and Water Treatment, Yokohama*, International Desalination Association, Topsfield, Mass., 1993; *Proceedings, IDA World Congress on Desalination and Water Sciences, Abu Dhabi*, International Desalination Association, Topsfield, Mass., 1995.
31. *Int. Desalin. Water Reuse Q.*, Lineal Publishing Co., Boca Raton, Fla.
32. A. Delyanis and E. Delyanis, eds., *Proceedings, Congress by Working Party on Fresh Water from the Sea of EFCE*, 1962, 1967, 1970, 1976, 1978, 1980; *Proceedings, European Federation of Chemical Engineers 12th International Symposium on Desalination and Water Reuse, Malta*, 1991. Several of the IDA congresses listed below (24) were also co-sponsored by the EFCE.
33. *Exodus* **15**: 22–25.
34. M. Balaban, *The Desalination Directory*, Balaban Research Publications, Mario Negri Sud Research Institute, Santa Maria Imbaro (Ch), Italy.
35. S. P. Kasper and N. Lior, *Desalination* **30**, 541 (1979).
36. L. A. Bromley, D. Singh, P. Ray, S. Sridhar, and S. M. Read, *AIChE J.* **20**, 326 (1974).
37. F. S. Aschner, in K. S. Spiegler and A. D. K. Laird, eds., *Principles of Desalination*, 2d ed., Pt. A, Academic Press, New York, 1980, Chapt. 5, p. 193.
38. F. Lennox and T. R. Stauffer, in Ref. 17, p. 27.
39. P. F. George, J. A. Manning, and C. F. Schrieber, *Desalination Materials Manual*, U.S. Department of the Interior, Office of Saline Water, Washington, D.C., 1975.
40. J. Glater, J. L. York, and K. S. Campbell, in K. S. Spiegler and A. D. K. Laird, eds., *Principles of Desalination*, 2nd ed., Pt. B, Academic Press, New York, 1980, Chapt. 10, p. 627.
41. H.-G. Heitman, *Saline Water Processing*, VCH Publications Inc., Weinheim, Germany, and New York, 1990.
42. V. B. Chernozubov and co-workers, in *Proceedings of the 1st International Symposium on Water Desalination*, Washington, D.C., Oct. 1965, p. 139.
43. CIBA-Geigy Corp., *The Application of Belgard EV and Belgard EVN in Sea Water Evaporators*, CIBA-GEIGY pamphlet DB 21.
44. B. M. Fabuss, in Ref. 38, Pt. B, Appndx. 2, p. 765.
45. F. M. A. Al-Awadi, *Int. Desalin. Water Reuse Q.* **4**(1), 16 (1991).
46. T. G. Temperley, *Int. Desalin. Water Reuse Q.* **4**(1), 12 (1991).
47. K. S. Spiegler and A. D. K. Laird, eds., *Principles of Desalination*, 2nd ed., Academic Press, New York, 1980.
48. A. S. Khan, *Desalination Processes and Multistage Flash Distillation Practice*, Elsevier, Amsterdam, the Netherlands, 1986.
49. N. Lior, ed., *Measurements and Control in Water Desalination*, Elsevier, Amsterdam, the Netherlands, 1986.
50. K. S. Spiegler and Y. M. El Sayed, *A Desalination Primer*, Balaban Desalination Publications, Mario Negri Sud Research Institute, Santa Maria Imbaro (Ch), Italy, 1994.
51. M. W. Kellogg Co., *Saline Water Data Conversion Engineering Data Book*, 3rd ed., Office of Saline Water Contract No. 14-30-2639, U.S. Department of the Interior, Washington, D.C., 1975.
52. N. Lior and R. Greif, *Desalination* **33**, 269 (1980).
53. O. Miyatake, T. Hashimoto, and N. Lior, *Int. J. Heat Mass Transfer* **35**, 3245 (1992).
54. O. Miyatake, T. Hashimoto, and N. Lior, *Desalination* **91**, 51 (1993).

55. Y. M. El-Sayed and R. S. Silver, in Ref. 38, Pt. A, Chapt. 2, p. 55.

56. A. B. Steinbruchel and R. D. Rhinesmith, in Ref. 38, Chapt. 3, p. 111.

57. Burns and Roe, Inc., *Universal Design—Report and User's Manual on Design of 2.5 Million Gallon per Day Universal Desalting Plant*, Vol. I–V, O.S.W. Contract No. 14-01-0001-955, U.S. Dept. Interior, Washington, D.C., 1969.

58. A. Ophir, *Desalination* **84**, 85 (1991).

59. F. C. Standiford, in Ref. 47, Chapt. 2.2, p. 263.

60. F. C. Standiford, in Ref. 47, Chapt. 2.3, p. 293.

61. **U.S. Pat. 2,821,304** (1958), A. Zarchin.

62. B. W. Tleimat, in Ref. 38, Pt. B, Chapt. 7, p. 359.

63. L. Dresner and J. S. Johnson, Jr., in Ref. 38, Pt. B, Chapt. 8, p. 401.

64. P. Hoornaert, *Reverse Osmosis*, Pergamon Press, New York, 1984.

65. S. Sourirajan and T. Matsuura, eds., *Reverse Osmosis and Ultrafiltration*, ACS Symposium Series 281, American Chemical Society, Washington, D.C., 1985.

66. D. Bhattacharya and M. E. Williams, in W. S. Winston Ho and K. K. Sirkar, eds., *Membrane Handbook*, Van Nostrand Reinhold, New York, 1992.

67. Z. Amjad, ed., *Reverse Osmosis: Membrane Technology, Water Chemistry and Industrial Applications*, Van Nostrand Reinhold, New York, 1993.

68. L. H. Shaffer and M. S. Mintz, in Ref. 38, Pt. A, Chapt. 6, p. 257.

69. S. B. Applebaum, *Demineralization by Ion Exchange*, Academic Press, New York, 1968.

70. D. Hoffman, *Desalination* **89**, 115 (1992).

71. S. E. Aly, *Wörme Stoffübertrag.* **20**, 263 (1986).

72. B. Doron, J. Weinberg, and U. Fisher, *IDA World Conference on Desalination and Water Re-Use*, Washington, D.C., 1991.

73. P. Glueckstern, *Desalination* **101**, 11 (1995).

74. A. Rabl, *Active Solar Collectors and Their Applications*, Oxford University Press, New York, 1985.

75. N. Lior, in F. de Winter, ed., *Collectors, Energy Storage, and Materials*, MIT Press, Cambridge, Mass., 1991, Chapt. 4, p. 99.

76. C. E. Nielsen, in W. C. Dickinson and P. N. Cheremisinoff, *Solar Energy Technology Handbook*, Part A, Marcel Dekker, New York, 1980, Chapt. 10, p. 345.

77. N. Lior, in G. O. G. Luf, ed., *Active Solar Systems*, MIT Press, Cambridge, Mass., 1993, Chapt. 17, p. 615.

78. M. Rey and F. Lauro, *Desalination* **39**, 159 (1981).

79. W. H. Avery and C. Wu, *Renewable Energy from the Ocean*, Oxford University Press, Oxford, 1994.

80. A. Maurel, *Desalination* **31**, 489 (1979).

81. W. Boesch, *Desalination* **41**, 233 (1982).

82. P. Gaiser and U. Plantikow, in Ref. 17, p. 411.

Noam Lior
University of Pennsylvania

Robert Bakish
Fairleigh Dickinson University

INDUSTRIAL WATER TREATMENT

Abundant supplies of fresh water are essential to the development of industry. Enormous quantities are required for the cooling of products and equipment, for process needs, for boiler feed, and for sanitary and potable water supply.

Water Impurities

Water impurities include dissolved and suspended solids. Calcium bicarbonate is a soluble salt. A solution of calcium bicarbonate is clear, because the calcium and bicarbonate are present as atomic-size ions that are not large enough to reflect light. Suspended solids are substances that are not completely soluble in water and are present as particles. These particles usually impart a visible turbidity to the water.

Clarification

Suspended matter in raw water supplies is removed by various methods to provide a water suitable for domestic purposes and most industrial requirements. The suspended matter can consist of large solids, settleable by gravity alone without any external aids, and nonsettleable material, often colloidal in nature. Removal is generally accomplished by coagulation, flocculation, and sedimentation. The combination of these three processes is referred to as conventional clarification.

Coagulation involves neutralizing charged particles to destabilize suspended solids. In most clarification processes, a flocculation step then follows. Flocculation starts when neutralized or entrapped particles begin to collide and fuse to form larger particles. This process can occur naturally or can be enhanced by the addition of polymeric flocculant aids.

Table 1 lists a number of common inorganic coagulants. Typical iron and aluminum coagulants are acid salts that lower the pH of the treated water by hydrolysis. Depending on initial raw water alkalinity and pH, an alkali such as lime or caustic must be added to counteract the pH depression of the primary coagulant. Iron and aluminum hydrolysis products play a significant role in the coagulation process, especially in cases in which low turbidity influent waters benefit from the presence of additional collision surface areas.

Table 1. Common Inorganic Coagulants

Name	Typical formula	Typical strength	Typical forms used in water treatment	Density	Typical uses
aluminum sulfate	Al ₂ (SO ₄) ₃ ·14 to 18H ₂ O	17° o Al ₂ O ₃	lump, granular, or powder	961–1121 kg m ³	primary coagulant
alum		8.25° o Al ₂ O ₃	liquid	1.33 kg L	
aluminum chloride	AlCl ₃ ·6H ₂ O	35° o AlCl ₃	liquid	1.50 kg L	primary coagulant
ferric sulfate	Fe ₂ (SO ₄) ₃ ·9H ₂ O	68° o Fe ₂ (SO ₄) ₃	granular	1121–1153 kg m ³	primary coagulant
ferric-floc	Fe ₂ (SO ₄) ₃ ·5H ₂ O	41° o Fe ₂ (SO ₄) ₃	solution	1.47 kg L	primary coagulant
ferric chloride	FeCl ₃	60° o FeCl ₃ , 35–45° o FeCl ₃	crystal, solution	961–1025 kg m ³ 1.34–1.49 kg L	primary coagulant
sodium aluminate	Na ₂ Al ₂ O ₄	38–46° o Na ₂ Al ₂ O ₄	liquid	1.47–1.55 kg L	primary coagulant; cold-hot precipi-tation softening

With aluminum sulfate, optimum coagulation efficiency and minimum floc solubility normally occur at pH 6.0–7.0. Iron coagulants can be used successfully over the much broader pH range of 5.0–11.0. If ferrous compounds are used, oxidation to ferric iron is needed for complete precipitation. This may require either chlorine addition or pH adjustment.

Polyelectrolytes refers to all water-soluble organic polymers used for clarification, whether they function as coagulants or flocculants. Water-soluble polymers may be classified as follows:

- Anionic. Ionize in water solution to form negatively charged sites along the polymer chain.
- Cationic. Ionize in water solution to form positively charged sites along the polymer chain.
- Nonionic. Ionize in water solution to form slight negatively charged sites along the polymer chain.

Polymeric primary coagulants are cationic materials with relatively low molecular weights (under 500,000). The cationic charge density (available positively charged sites) is high. Polymeric flocculants or coagulant aids may be anionic, cationic, or nonionic. Their molecular weights may be as high as 50,000,000. Table 2 describes some typical organic polyelectrolytes.

Table 2. Common Organic Polyelectrolytes

Polymer type	Typical formula	Typical molecular weight	Available forms	Typical uses
nonionic	polyacrylamide $\left[\begin{array}{c} \text{—CH}_2\text{—CH—} \\ \\ \text{C=O} \\ \\ \text{NH}_2 \end{array} \right]_n$	1 × 10 ⁶ to 2 × 10 ⁷	powder, emulsion, solution	flocculant in clarification with inorganic or organic coagulants

anionic	hydrolyzed polyacrylamide $\left[\begin{array}{c} \text{CH}_2 - \text{CH} \\ \\ \text{C}=\text{O} \\ \\ \text{NH}_2 \end{array} \right]_n \left[\begin{array}{c} \text{CH}_2 - \text{CH} \\ \\ \text{C}=\text{O} \\ \\ \text{ONa} \end{array} \right]_y$	1×10^6 to 2×10^7	powder, emulsion, solution	floculant in clarification with inorganic or organic coagulants
cationic	poly(DADMAC) or poly(DMDAAC) polymers $\left[\begin{array}{c} \text{CH}_2 - \text{CH} - \text{CH} - \text{CH}_2 \\ \qquad \\ \text{CH}_2 \quad \text{CH}_2 \\ \diagdown \quad / \\ \text{N} \\ / \quad \backslash \\ \text{CH}_3 \quad \text{CH}_3 \end{array} \right]_n + \text{Cl}^-$	250 to 500×10^3	solution	primary coagulant alone or in combina-tion with inorganics in clarification
cationic	quarternized polyamines $\left[\begin{array}{c} \text{CH}_2 - \text{CH} - \text{CH}_2 - \text{N} - \text{CH}_3 \\ \qquad \qquad \\ \text{OH} \quad \text{CH}_3 \end{array} \right]_n$	10 to 500×10^4	solution	primary coagulant alone or in combina-tion with inorganics in clarification
cationic	polyamines $\left[\begin{array}{c} \text{CH}_2 - \text{CH} - \text{CH}_2 - \text{NH}_2 \\ \qquad \\ \text{OH} \quad \text{CH}_3 \end{array} \right]_n$	10^4 to 10	solution	primary coagulant alone or in combina-tion with inorganics in clarification

The use of organic polymers offers several advantages over the use of inorganic coagulants:

The amount of sludge produced during clarification can be reduced by 50–90%; the approximate dry weights of solids removed per pound of dry alum and ferric sulfate are approximately 0.11 and 0.23 kg, respectively.

The resulting sludge contains less chemically bound water and can be more easily dewatered.

Polymeric coagulants do not affect pH; therefore, the need for supplemental alkalinity, such as lime, caustic, or soda ash, is reduced or eliminated.

Polymeric coagulants do not add to the total dissolved solids concentration, eg, 1 ppm of alum adds 0.45 ppm of sulfate ion (expressed as CaCO₃); the reduction in sulfate can significantly extend the capacity of anion-exchange systems.

Soluble iron or aluminum carryover in the clarifier effluent may result from inorganic coagulant use; therefore, elimination of the inorganic coagulant can minimize the deposition of these metals in filters, ion-exchange units, and cooling systems.

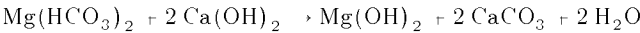
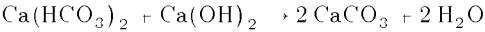
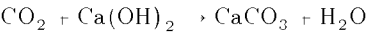
In certain instances, an excess of primary coagulant (whether inorganic, polymeric, or a combination of both) may be fed to promote large floc size and to increase settling rate. However, in some waters, even high doses of primary coagulant will not produce the desired effluent clarity. A polymeric coagulant aid added after the primary coagulant may, by developing a larger floc at low treatment levels, reduce the amount of primary coagulant required.

Generally, very high molecular weight anionic polyacrylamides are the most effective coagulant aids. Nonionic or cationic types have proven successful in some clarifier systems. Essentially, the polymer bridges the small floc particles and causes them to agglomerate rapidly into larger, more cohesive flocs that settle quickly. The higher molecular weight polymers bridge suspended solids most effectively. Coagulant aids have proven quite successful in precipitation softening and clarification to achieve improved settling rates of precipitates and finished water clarity.

Precipitation Softening

Precipitation softening processes are used to reduce raw water hardness, alkalinity, silica, and other constituents. This helps prepare water for direct use as cooling tower makeup or as a first-stage treatment followed by ion exchange for boiler makeup or process use. The water is treated with lime or a combination of lime and soda ash (carbonate ion). These chemicals react with the hardness and natural alkalinity in the water to form insoluble compounds. The compounds precipitate and are removed from the water by sedimentation and, usually, filtration. Waters with moderate to high hardness and alkalinity concentrations (150–500 ppm as CaCO₃) are often treated in this fashion.

Cold Lime Softening. Precipitation softening accomplished at ambient temperatures is referred to as cold lime softening. When hydrated lime, Ca(OH)₂, is added to the water being treated, the following reactions occur:



Noncarbonate or permanent calcium hardness, if present, is not affected by treatment with lime alone. If noncarbonate magnesium hardness is present in an amount greater than 70 ppm and an excess hydroxyl alkalinity of about 5 ppm is maintained, the magnesium will be reduced to about 70 ppm, but the calcium will increase in proportion to the magnesium reduction.

To improve magnesium reduction, which also improves silica reduction in cold process softening, sodium aluminate may be used. The sodium aluminate provides hydroxyl ion (OH⁻) needed for improved magnesium reduction, without increasing calcium hardness in the treated water. In addition, the hydrolysis of sodium aluminate results in the formation of aluminum hydroxide, which aids in floc formation, sludge blanket conditioning, and silica reduction.

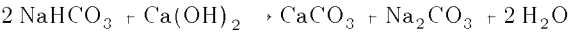
Warm Lime Softening. The warm lime softening process operates in the temperature range of 49–60°C. The solubilities of calcium, magnesium, and silica are reduced by increased temperature. Therefore, they are more effectively removed by warm lime softening than by cold lime softening.

Hot Process Softening. Hot process softening is usually carried out under pressure at temperatures of 108–116°C. At the operating temperature, hot process softening reactions go essentially to completion. This treatment method involves the same reactions described above, except that raw water CO₂ is vented and does not participate in the lime reaction. The use of lime and soda ash permits hardness reduction down to 0.5 g/gal, or about 8 ppm, as calcium carbonate. Magnesium is reduced to 2–5 ppm, because of the lower solubility of magnesium hydroxide at the elevated temperatures.

Silica Reduction. Hot process softening can also provide good silica reduction. The silica reduction is accomplished through adsorption of the silica on the magnesium hydroxide precipitate. If there is insufficient magnesium present in the raw water to reduce silica to the desired level, magnesium compounds (such as magnesium oxide, magnesium sulfate, magnesium carbonate, and dolomitic lime) may be used.

Alkalinity Reduction. Treatment by lime precipitation reduces alkalinity. However, if the raw water alkalinity exceeds the total hardness, sodium bicarbonate alkalinity is present. In such cases, it is usually necessary to reduce treated water alkalinity in order to reduce condensate system corrosion or permit increased cycles of concentration.

Treated Water Quality. Predicted analyses of a typical raw water treated by various lime and lime–soda softening processes are presented in Table 3. Treatment by lime converts the sodium bicarbonate in the raw water to sodium carbonate as follows:



Calcium sulfate (gypsum) may be added to reduce the carbonate to required levels. The reaction is as follows:



Table 3. Typical Softener Effluent Analyses

Factor	Raw water	Removal of calcium alkalinity cold-lime	Lime–soda softening (cold)	Lime–soda softening (hot) ^a	Lime softening (hot) ^a
total hardness (as CaCO ₃), ppm	250	145	81	20	120
calcium hardness (as CaCO ₃), ppm	150	85	35	15	115
magnesium hardness (as CaCO ₃), ppm	100	60	46	5	5
"P" alkalinity (as CaCO ₃), ppm	0	27	37	23	18
"M" alkalinity (as CaCO ₃), ppm	150	44	55	40	28
silica (as SiO ₂), ppm	20	19	18	1–2	1–2
pH	7.5	10.3	10.6	10.5	10.4

^a Removal of SiO₂ by the hot process, to the levels shown, may require the feed of supplemental magnesium oxide. Sludge recirculation is necessary. All raw water constituents will be diluted by the steam used for heating by approximately 15% if the process is hot.

Filtration

Filtration is used in addition to regular coagulation and sedimentation or precipitation softening for removal of solids from surface water or waste water. This prepares the water for use as potable, boiler, or cooling makeup. Waste water filtration helps users meet more stringent effluent discharge permit requirements.

Filtration does not remove dissolved solids, but may be used together with a softening process, which does reduce the concentration of dissolved solids. For example, anthracite filtration is used to remove residual precipitated hardness salts remaining after precipitation softening.

In most water clarification or softening processes in which coagulation and precipitation occur, at least a portion of the clarified water is filtered. Clarifier effluents of 2–10 NTU may be improved to 0.1–1.0 NTU by conventional sand filtration. Filtration ensures acceptable suspended solids concentrations in the finished water even when upsets occur in the clarification processes.

Ion Exchange

Ion exchangers exchange one ion for another, hold it temporarily, and then release it to a regenerant solution. In an ion-exchange system, undesirable ions in the water supply are replaced with more acceptable ions.

Ionizable groups attached to the resin bead determine the functional capability of the resin. Industrial water treatment resins are classified into four basic categories: strong acid cation (SAC); weak acid cation (WAC); strong base anion (SBA); and weak base anion (WBA).

SAC resins can neutralize strong bases and convert neutral salts into their corresponding acids. SBA resins can neutralize strong acids and convert neutral salts into their corresponding bases. These resins are used in most softening and full demineralization applications. WAC and WBA resins are able to neutralize strong bases and acids, respectively. These resins are used for dealkalization, partial demineralization, or (in combination with strong resins) full demineralization.

Sodium Zeolite Softening. Sodium zeolite softening is the most widely applied use of ion exchange. In zeolite softening, water containing scale-forming ions, such as calcium and magnesium, passes through a resin bed containing SAC resin in the sodium form. In the resin, the hardness ions are exchanged with the sodium, and the sodium diffuses into the bulk water solution. The hardness-free water, termed soft water, can then be used for low to medium pressure boiler feedwater, reverse osmosis system makeup, some chemical processes, and commercial applications, such as laundries.

Demineralization. Softening alone is insufficient for most high-pressure boiler feed waters and for many process streams, especially those used in the manufacture of electronics equipment. In addition to the removal of hardness, these processes require removal of all dissolved solids, such as sodium, silica, alkalinity, and the mineral anions (Cl⁻, SO₄²⁻, and NO₃⁻).

Demineralization of water is the removal of essentially all inorganic salts by ion exchange. In this process, strong acid cation resin in the hydrogen form converts dissolved salts into their corresponding acids, and strong base anion resin in the hydroxide form removes these acids. Demineralization produces water similar in quality to distillation at a lower cost for most fresh waters.

The standard cation–anion process has been modified in many systems to reduce the use of costly regenerants and the production of waste. Modifications include the use of decarbonators, weak acid and weak base resins. Several different approaches to demineralization using these processes are shown in Figure 1.

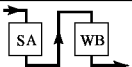
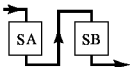
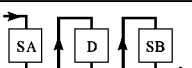
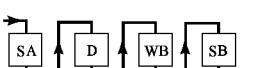
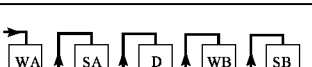
System	Application	Typical Effluent	Advantages and Limitations
	Silica and CO ₂ are not objectionable	Conductance: 10-40 intro silica unchanged	Low equipment costs low regenerant costs
	Lower alkalinity, raw water, silica, and CO ₂ removal required	Conductance: <15 intro silica: 0.02-0.10 ppm	Low equipment costs Medium regenerant costs
	High alkalinity, raw water, silica, and CO ₂ removal required	Conductance: <15 intro silica: 0.02-0.10 ppm	Low regenerant costs Repumping required
	High alkalinity, chloride and sulfate, raw water, silica, and CO ₂ removal required	Conductance: <15 intro silica: 0.02-0.10 ppm	Higher equipment cost lowest regenerant cost repumping required
	High hardness, alkalinity, chloride, and sulfate, raw water, silica, and CO ₂ removal required	Conductance: <10 intro silica: 0.02-0.06 ppm	Higher equipment cost lowest regenerant cost repumping required
	High sodium, raw water, low leakage required	Conductance: <10 intro silica: 0.02-0.06 ppm	Medium equipment cost lower acid cost for leakage obtained
	High sodium, raw water, existing 2-bed system, low leakage required	Conductance: <5 intro silica: 0.02-0.06 ppm	Easy to retrofit system danger of acidic water on anion breakthrough
	Low solids, raw water, high purity required	Conductance: <1 intro silica: 0.01-0.05 ppm	Low equipment cost high chemical cost, higher attention required
	High solids, water, high purity required	Conductance: <1 intro silica: 0.01-0.05 ppm	Medium equipment cost high chemical cost, higher attention required

Fig. 1. Demineralizer systems consist of various unit processes arranged to meet the system needs. [SA] Strong acid cation exchanger; [SB] Strong base anion exchanger; [D] Degasifier; [MB] Mixed bed; [WA] Weak acid cation exchanger; [WB] Weak base anion exchanger; and [CF] Counterflow cation.

Condensate Polishing. Ion exchange can be used to purify or polish returned condensate, removing corrosion products that could cause harmful deposits in boilers. Typically, the contaminants in the condensate system are particulate iron and copper. Low levels of other contaminants may enter the system through condenser and pump seal leaks or carryover of boiler water into the steam. Condensate polishers filter out the particulates and remove soluble contaminants by ion exchange.

Most paper mill condensate polishers operate at temperatures approaching 93°C, precluding the use of anion resin. Cation resin, which is stable up to temperatures of over 132°C, is used for deep bed condensate polishing in these applications. The resin is regenerated with sodium chloride brine, as in a zeolite softener. In situations in which sodium leakage from the polisher adversely affects the boiler water internal chemical program or steam tempering water purity, the resin can be regenerated with an ionized amine solution to prevent these problems.

Membrane Processes

In recent years, membrane processes have been used increasingly for the production of "pure" waters from fresh water and seawater. Membrane processes are also being applied in process and wastewater systems. Although often thought to be expensive and relatively experimental, membrane technology is advancing quickly, becoming less expensive, improving performance, and extending life expectancy.

Common membrane processes include ultrafiltration (UF), reverse osmosis (RO), electrodialysis (ED), and electrodialysis reversal (EDR). These processes (with the exception of UF) remove most ions; RO and UF systems also provide efficient removal of nonionized organics and particulates. Because UF membrane porosity is too large for ion rejection, the UF process is used to remove contaminants, such as oil and grease, and suspended solids.

Reverse Osmosis. Osmosis is the flow of solvent through a semipermeable membrane, from a dilute solution to a concentrated solution. This flow results from the driving force created by the difference in pressure between the two solutions. Osmotic pressure is the pressure that must be added to the concentrated solution side to stop the solvent flow through the membrane. Reverse osmosis is the process of reversing the flow, forcing water through a membrane from a concentrated solution to a dilute solution to produce pure water. Figure 2 illustrates the processes of osmosis and reverse osmosis.

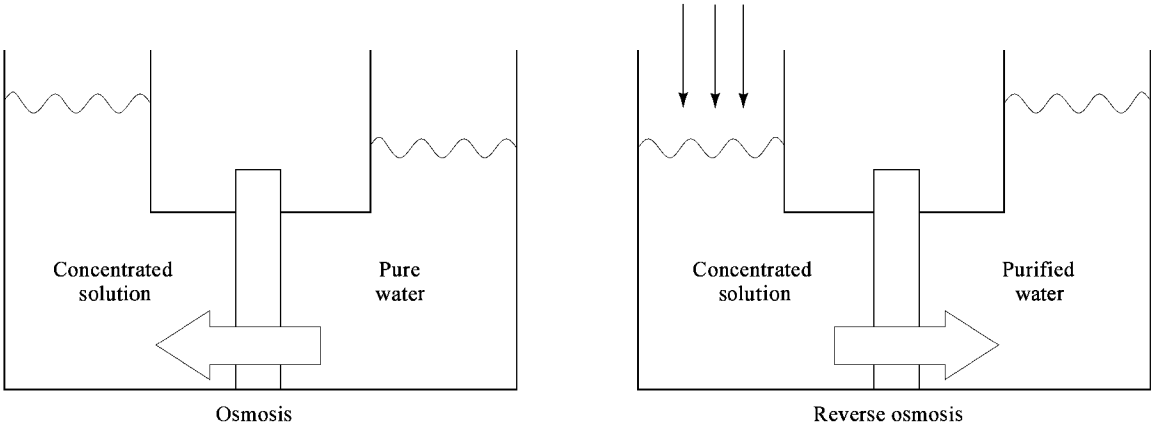


Fig. 2. In the osmosis process, water flows through a membrane from a dilute solution to a more concentrated solution. In reverse osmosis, applied pressure causes water to flow in the opposite direction.

Reverse osmosis is created when sufficient pressure is applied to the concentrated solution to overcome the osmotic pressure. This pressure is

provided by feed-water pumps. Concentrated contaminants (brine) are removed from the high pressure side of the RO membrane, and pure water (permeate) is removed from the low pressure side. Membrane modules may be staged in various design configurations, producing the highest quality permeate with the least amount of waste.

Typically, 95% of dissolved salts are removed from the brine. All particulates are removed. However, due to their molecular porosity, RO membranes do not remove dissolved gases, such as Cl_2 , CO_2 , and O_2 .

Electrodialysis. Electrodialysis processes transfer ions of dissolved salts across membranes, leaving purified water behind. Ion movement is induced by direct current electrical fields. A negative electrode (cathode) attracts cations, and a positive electrode (anode) attracts anions. Systems are compartmentalized in stacks by alternating cation and anion transfer membranes. Alternating compartments carry concentrated brine and purified permeate. Typically, 40–60% of dissolved ions are removed or rejected. Further improvement in water quality is obtained by staging (operation of stacks in series). ED processes do not remove particulate contaminants or weakly ionized contaminants, such as silica.

Electrodialysis Reversal. Electrodialysis reversal processes operate on the same principles as ED; however, EDR operation reverses system polarity (typically three to four times per hour). This reversal stops the buildup of concentrated solutions on the membrane and thereby reduces the accumulation of inorganic and organic deposition on the membrane surface. EDR systems are similar to ED systems, designed with adequate chamber area to collect both product water and brine. EDR produces water of the same purity as ED.

Boilers

Boiler System Corrosion. The dissolved gases normally present in water cause many corrosion problems. For instance, oxygen in water produces pitting that is particularly severe because of its localized nature (Fig. 3). Carbon dioxide corrosion is frequently encountered in condensate systems and less commonly in water distribution systems. Water containing ammonia, particularly in the presence of oxygen, readily attacks copper and copper-bearing alloys. The resulting corrosion leads to deposits on boiler heat transfer surfaces and reduces efficiency and reliability. In addition to the gases, corrosion can also be caused by the concentration of caustic or acidic species in the feed water.



Fig. 3. Tube severely damaged by oxygen.

Caustic Corrosion. Concentration of caustic (NaOH) can occur either as a result of steam blanketing (which allows salts to concentrate on boiler metal surfaces) or by localized boiling beneath porous deposits on tube surfaces.

Caustic corrosion (gouging) occurs when caustic is concentrated and dissolves the protective magnetite (Fe_3O_4) layer. Iron, in contact with the boiler water, forms magnetite and the protective layer is continuously restored. However, as long as a high caustic concentration exists, the magnetite is constantly dissolved, causing a loss of base metal and eventual failure (Fig. 4).



Fig. 4. Caustic gouging caused failure of this boiler tube.

Steam blanketing is a condition that occurs when a steam layer forms between the boiler water and the tube wall. Under this condition, insufficient water reaches the tube surface for efficient heat transfer. The water that does reach the overheated boiler wall is rapidly vaporized, leaving behind a concentrated caustic solution, which is corrosive.

Porous metal oxide deposits also permit the development of high boiler water concentrations. Water flows into the deposit and heat applied to the tube causes the water to evaporate, leaving a concentrated solution. Again, corrosion may occur. Caustic attack creates irregular patterns, often referred to as gouges. Deposition may or may not be found in the affected area.

Acidic Corrosion. Low makeup or feed-water pH can cause serious acid attack on metal surfaces in the preboiler and boiler system. Even if the original makeup or feed-water pH is not low, feed water can become acidic from contamination of the system. Common causes include improper operation or control of demineralizer cation units, process contamination of condensate (eg, sugar contamination in food processing plants), and cooling water contamination from condensers. Acid corrosion can also be caused by chemical cleaning operations. Overheating of the cleaning solution can cause breakdown of the inhibitor used, excessive exposure of metal to cleaning agent, and high cleaning agent concentration. Failure to neutralize acid solvents

completely before startup has also caused problems. In a boiler and feed-water system, acidic attack can take the form of general thinning, or it can be localized at areas of high stress such as drum baffles, U bolts, acorn nuts, and tube ends.

Hydrogen Embrittlement. Hydrogen embrittlement is rarely encountered in industrial plants. The problem usually occurs only in boilers operating at or above 689.5×10^3 Pa. Hydrogen embrittlement of mild steel boiler tubing occurs in high pressure boilers when atomic hydrogen forms at the boiler tube surface as a result of corrosion. Hydrogen permeates the tube metal, where it can react with iron carbides to form methane gas, or with other hydrogen atoms to form hydrogen gas. These gases evolve predominantly along grain boundaries of the metal. The resulting increase in pressure leads to metal failure.

The initial surface corrosion that produces hydrogen usually occurs beneath a hard, dense scale. Acidic contamination or localized low pH excursions are normally required to generate atomic hydrogen. In high purity systems, raw water in-leakage (eg, condenser leakage) lowers boiler water pH when magnesium hydroxide precipitates, resulting in corrosion, formation of atomic hydrogen, and initiation of hydrogen attack.

Stress Corrosion Cracking. Stress corrosion cracking occurs from the combined action of corrosion and stress. The corrosion may be initiated by improper chemical cleaning, high dissolved oxygen levels, pH excursions in the boiler water, the presence of free hydroxide, and high levels of chlorides. Stresses are either residual in the metal or caused by thermal excursions. Rapid startup or shutdown can cause or further aggravate stresses. Tube failures occur near stressed areas such as welds, supports, or cold worked areas.

Boiler Deposits. Deposition is a principal problem in the operation of steam generating equipment. The accumulation of material on boiler surfaces can cause overheating and or corrosion. Both of these conditions frequently result in unscheduled downtime. Common feed-water contaminants that can form boiler deposits include calcium, magnesium, iron, copper, aluminum, silica, and (to a lesser extent) silt and oil. Most deposits can be classified as one of two types: scale that crystallized directly onto tube surfaces or sludge deposits that precipitated elsewhere and were transported to the metal surface by the flowing water.

Boiler feed water pretreatment systems have advanced to such an extent that it is now possible to provide boilers with ultrapure water. However, this degree of purification requires the use of elaborate pretreatment systems. The capital expenditures for such pretreatment equipment trains can be considerable and are often not justified when balanced against the capability of internal treatment.

The quality of feed water required depends on boiler operating pressure, design, heat transfer rates, and steam use. Most boiler systems have sodium zeolite softened or demineralized makeup water. Feed-water hardness usually ranges from 0.01 to 2.0 ppm, but even water of this purity does not provide deposit-free operation. Therefore, good internal boiler water treatment programs are necessary.

Boiler Water Treatment.

Oxygen Control. To meet industrial standards for both oxygen content and the allowable metal oxide levels in feed water, nearly complete oxygen removal is required. This can be accomplished only by efficient mechanical deaeration supplemented by an effective and properly controlled chemical oxygen scavenger.

To deaerate the boiler feedwater, water is sprayed into a steam atmosphere. This heats the water to within a few degrees of the temperature of the saturated steam. Because the solubility of oxygen in water is low under these conditions, 97–98% of the oxygen in the incoming water is released to the steam and is purged from the system by venting. Although the remaining oxygen is not soluble under equilibrium conditions, it is not readily released to the steam. Therefore, water leaving the heating section of the deaerator must be scrubbed vigorously with steam to maximize removal.

In addition to mechanical deaeration, chemical oxygen scavengers are used to remove any remaining oxygen. The oxygen scavengers most commonly used in boiler systems are sodium sulfite, sodium bisulfite, hydrazine, catalyzed versions of the sulfites and hydrazine, and organic oxygen scavengers, such as hydroquinone and ascorbate.

It is of critical importance to select and properly use the best chemical oxygen scavenger for a given system. Principal factors that determine the best oxygen scavenger for a particular application include reaction speed, residence time in the system, operating temperature and pressure, and feed-water pH. Interferences with the scavenger–oxygen reaction, decomposition products, and reactions with metals in the system are also important factors. Other contributing factors include the use of feed water for attemperation, the presence of economizers in the system, and the end use of the steam. Chemical oxygen scavengers should be fed to allow ample time for the scavenger–oxygen reaction to occur. The deaerator storage system and the feed-water storage tank are commonly used feed points.

Scale and deposits are controlled through the use of phosphates, chelants, and polymers. Phosphates are precipitating treatments, and chelants are solubilizing treatments. Polymers are most widely used to disperse particulates but they are also used to solubilize contaminants under certain conditions.

Phosphate Treatment. Calcium phosphate is virtually insoluble in boiler water. Even small levels of phosphate can be maintained to ensure the precipitation of calcium phosphate in the bulk boiler water, away from heating surfaces. Therefore, the introduction of phosphate treatment eliminates the formation of calcium carbonate scale on tube surfaces. When calcium phosphate is formed in boiler water of sufficient alkalinity, a particle with a relatively nonadherent surface charge is produced. This does not prevent the development of deposit accumulations over time, but the deposits can be controlled reasonably well by blowdown.

In a phosphate precipitation treatment program, the magnesium portion of the hardness contamination is precipitated preferentially as magnesium silicate. If silica is not present, the magnesium will precipitate as magnesium hydroxide. If insufficient boiler water hydroxide is being maintained, magnesium can combine with phosphate. Magnesium phosphate has a surface charge that can cause it to adhere to tube surfaces and then collect other solids. For this reason, alkalinity is an important part of a phosphate precipitation program.

Phosphate–Polymer Control. Phosphate treatment results are improved by organic supplements. Naturally occurring organics such as lignins, tannins, and starches were the first supplements used. The organics were added to promote the formation of a fluid sludge that would settle in the mud drum. Bottom blowdown from the mud drum removed the sludge.

There have been many advances in organic treatments. Synthetic polymers are now used widely, and the emphasis is on dispersion of particles rather than fluid sludge formation. Although this mechanism is quite complex, polymers alter the surface area and the surface charge to mass ratio of the boiler solids. Many synthetic polymers are used in phosphate precipitation programs. Most are effective in dispersing magnesium silicate and magnesium hydroxide as well as calcium phosphate. The polymers are usually low in molecular weight and have numerous active sites. Some polymers are used specifically for hardness salts or for iron; some are effective for a broad spectrum of ions.

Chelant Control. Chelants are the prime additives in a solubilizing boiler water treatment program. Chelants have the ability to complex many cations (hardness and heavy metals under boiler water conditions). They accomplish this by locking metals into a soluble organic ring structure. The chelated cations do not deposit in the boiler. When applied with a dispersant, chelants produce clean waterside surfaces.

Chelant–Polymer Control. Iron oxide is of particular concern in today's boiler water treatment programs. Deposition from low (less than 1.0 ppm) hardness boiler feed water is eliminated with chelant programs and can be reduced by up to 95% by a good polymer–phosphate treatment program. Iron oxide is an increasingly significant contributor to boiler deposits, because of the virtual elimination of hardness deposits in many systems and because the high heat transfer rates of many boilers encourage iron deposition.

A chelant–polymer combination is an effective approach to controlling iron oxide. Adequate chelant is fed to complex hardness and soluble iron, with a slight excess to solubilize iron contamination. Polymers are then added to condition and disperse any remaining iron oxide contamination.

A chelant–polymer program can produce clean waterside surfaces, contributing to much more reliable boiler operation. Out-of-service boiler cleaning schedules can be extended and, in some cases, eliminated. This depends on operational control and feed-water quality. Chelants with high complexing stabilities are "forgiving" treatments; they can remove deposits that form when feed-water quality or treatment control periodically deviates from standard.

Phosphate–Chelant–Polymer Combinations. Combinations of polymer, phosphate, and chelant are commonly used to produce results comparable to chelant–polymer treatment in boilers operating at 4137×10^3 Pa or less. Boiler cleanliness is improved over phosphate treatment, and the presence of phosphate provides an easy means of testing to confirm the presence of treatment in the boiler water.

Polymer-only Treatment. Polymer-only treatment programs are also used with a degree of success. In this treatment, the polymer is usually used as a weak chelant to complex the feed-water hardness. These treatments are most successful when feed-water hardness is consistently low.

High Pressure Boiler Water Treatment. High pressure boilers usually have feed water composed of demineralized makeup water and a high percentage of condensate returns. Because of these conditions, high pressure boilers are prone to caustic and acid attack. Low pressure boilers that use demineralized water and condensate as feed water are also susceptible to caustic and acid attack.

There are several means by which boiler water can become highly concentrated. One of the most common is iron oxide deposition on radiant wall tubes. Iron oxide deposits are often quite porous and act as miniature boilers. Water is drawn into the iron oxide deposit. Heat applied to the deposit from the tube wall generates steam, which passes out through the deposit. More water enters the deposit, taking the place of the steam. This cycle is repeated and the water beneath the deposit is concentrated to extremely high levels. It is possible to have 100,000 ppm of caustic beneath the deposit while the bulk water contains only about 5–10 ppm of caustic.

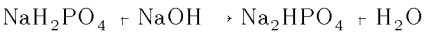
Boiler feed-water systems that use demineralized or evaporated makeup or pure condensate may be protected from caustic attack through coordinated phosphate and pH control. Phosphate buffers the boiler water, reducing the chance of large pH changes due to the development of high caustic or acid concentrations. Excess caustic combines with disodium phosphate and forms trisodium phosphate. Sufficient disodium phosphate must be available to combine with all of the free caustic in order to form trisodium phosphate.

Disodium phosphate neutralizes caustic by the following reaction:

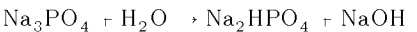


This results in the prevention of caustic buildup beneath deposits or within a crevice where leakage is occurring. Caustic corrosion (and caustic embrittlement, discussed later) does not occur, because high caustic concentrations do not develop.

Figure 5 shows the phosphate–pH relationship recommended to control boiler corrosion. Different forms of phosphate consume or add caustic as the phosphate shifts to the proper form. For example, addition of monosodium phosphate consumes caustic as it reacts with caustic to form disodium phosphate in the boiler water according to the following reaction:



Conversely, addition of trisodium phosphate adds caustic, increasing boiler water pH:



Control is achieved through feed of the proper type of phosphate either to raise or to lower the pH while maintaining the proper phosphate level. Increasing blowdown lowers both phosphate and pH. Therefore, various combinations and feedrates of phosphate, blowdown adjustment, and caustic addition are used to maintain proper phosphate–pH levels.

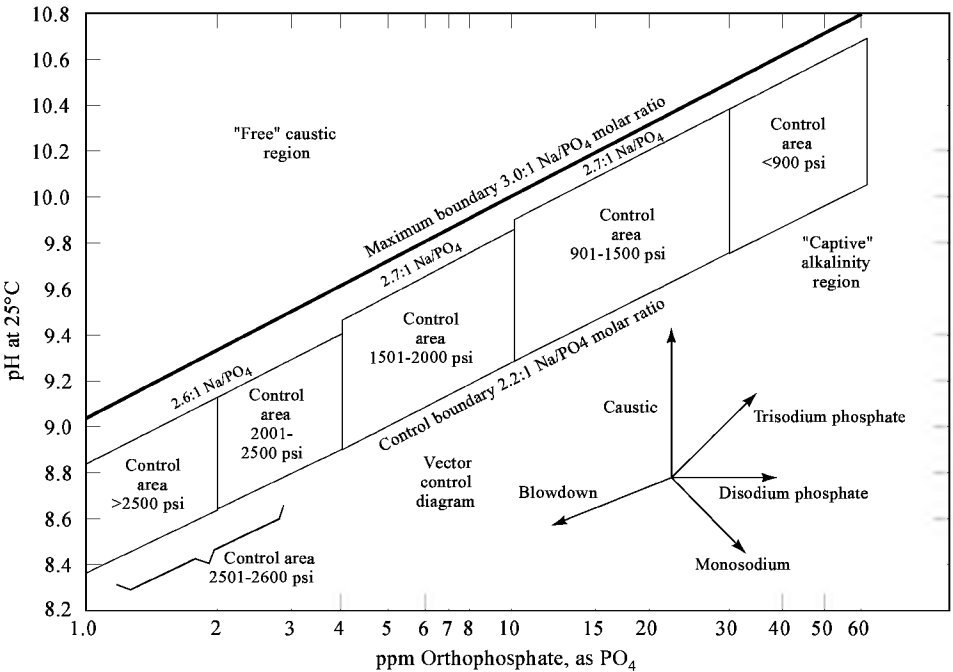


Fig. 5. Coordinated phosphate–pH control avoids both acid and caustic corrosion. To convert psi to kPa, divide by 6.895.

Elevated temperatures at the boiler tube wall or deposits can result in some precipitation of phosphate. This effect, termed *phosphate hideout*, usually occurs when loads increase. When the load is reduced, phosphate reappears.

Clean boiler water surfaces reduce potential concentration sites for caustic. Deposit control treatment programs, such as those based on chelants and synthetic polymers, can help provide clean surfaces.

When steam blanketing is occurring, corrosion can take place even without the presence of caustic, due to the steam–magnetite reaction and the dissolution of magnetite. In such cases, operational changes or design modifications may be necessary to eliminate the cause of the problem.

If deposits are minimized, the areas where caustic can be concentrated is reduced. To minimize the iron deposition in $6.895\text{--}12.07 \times 10^6$ Pa boilers, specific polymers have been designed to disperse the iron and keep it in the bulk water. As with phosphate precipitation and chelant control programs, the use of these polymers with coordinated phosphate–pH treatment improves deposit control.

Supercritical boilers use all-volatile treatments, generally consisting of ammonia and hydrazine. Because of the extreme potential for deposit formation and steam contamination, no solids can be tolerated in supercritical once-through boiler water, including treatment solids.

Steam Purity. Boiler water solids carried over with steam form deposits in nonreturn valves, superheaters, and turbine stop and control valves. Carryover can contaminate process streams and affect product quality. Deposition in superheaters can lead to failure due to overheating and corrosion, as shown in Figure 6.

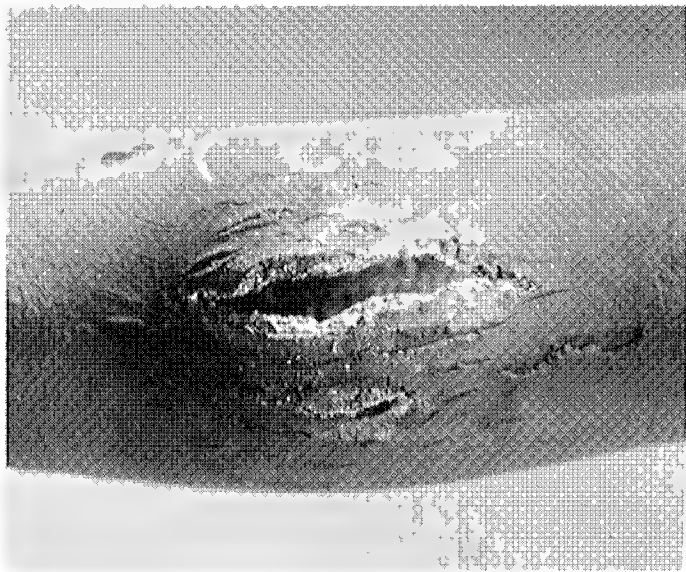


Fig. 6. Boiler water contamination of the steam caused superheater deposits, which led to tube metal overheating and failure.

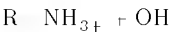
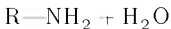
Superheated steam turbines are particularly prone to damage by carryover. Sticking of governor and stop valves due to deposits can cause turbine overspeed and catastrophic damage. Solid particles in steam can erode turbine parts, while deposition on turbine blades can reduce efficiency and capacity. Losses of 5° in turbine efficiency and 20° in turbine capacity have occurred due to deposition. When large slugs of boiler water carry over with steam, the resulting thermal and mechanical shock can cause severe damage.

Steam can be contaminated with solids even when carryover is not occurring. Contaminated spray attemperating water, used to control superheated steam temperature at the turbine inlet, can introduce solids into steam. A heat exchanger coil may be placed in the boiler mud drum to provide attemperation of the superheated steam. Because the mud drum is at a higher pressure than superheated steam, contamination will occur if leaks develop in the coil.

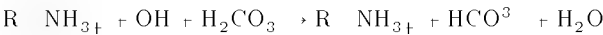
Chemical Treatment of Condensate Systems

Condensate systems can be chemically treated to reduce metal corrosion. Treatment chemicals include neutralizing amines, filming amines, and oxygen scavenger-metal passivators.

Neutralizing Amines. Neutralizing amines are used to neutralize the acid (H⁺) generated by the dissolution of carbon dioxide or other acidic process contaminants in the condensate. These amines hydrolyze when added to water and generate the hydroxide ions required for neutralization:



The overall neutralization reaction can be written as shown:



By regulating the neutralizing amine feed rate, the condensate pH can be elevated to within a desired range (eg, 8.8–9.2 for a mixed copper–iron condensate system).

Filming Amines. Another approach to controlling condensate system corrosion is the use of chemicals that form a protective film on metal surfaces (Fig. 7). This approach has come into widespread use with the development of suitable products containing long-chain nitrogenous materials. Filming amines protect against oxygen and carbon dioxide corrosion by replacing the loose oxide scale on metal surfaces with a thin amine film barrier.

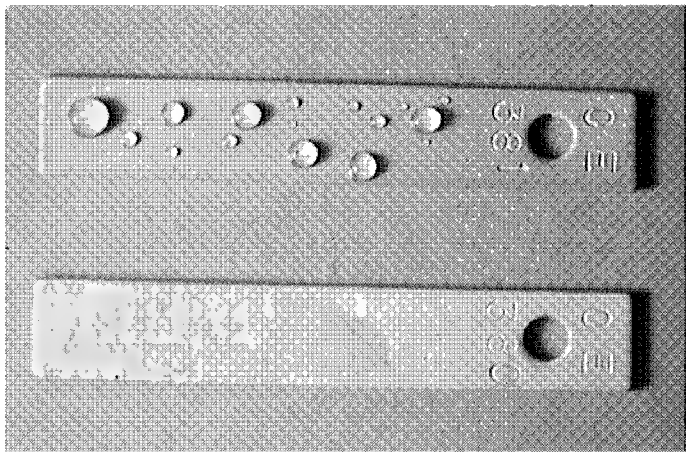


Fig. 7. Test specimens illustrate the nonwetttable surface produced by a filming amine (381) compared with an untreated surface (380).

Advances have been made in formulating filming amine treatments. Straight filming amines containing one ingredient, such as octadecylamine, are effective but often fail to cover the entire system and can produce fouling. Emulsifiers and, in some cases, small amounts of neutralizing amines can be added to improve film distribution by providing more uniform coverage. This increases system protection and reduces the fouling potential. Application experience has shown that combination amines (filming and neutralizing amines with dispersant aids) provide a superior film bond, reduce deposit problems, and provide better system coverage and thus provide more complete and economical corrosion protection.

Oxygen Scavenging and Metal Passivation. Where oxygen invades the condensate system, corrosion of iron and copper-bearing components can be overcome through proper pH control and the injection of an oxygen scavenger. One important factor to consider in choosing an oxygen scavenger for condensate treatment is its reactivity with oxygen at the temperature and pH of the system. A scavenger that removes oxygen rapidly

provides the best protection for the condensate metallurgy. Hydroquinone has been shown to be particularly effective for most systems.

The use of neutralizing amines in conjunction with an oxygen scavenger–metal passivator improves corrosion control in two ways. First, because any acidic species present is neutralized and pH is increased, the condensate becomes less corrosive. Second, most oxygen scavenger–passivators react more rapidly at the mildly alkaline conditions maintained by the amine than at lower pH levels. For these reasons, this combination treatment is gaining wide acceptance, particularly for the treatment of condensate systems that are contaminated by oxygen.

Cooling Systems

Cooling System Corrosion Corrosion can be defined as the destruction of a metal by chemical or electrochemical reaction with its environment. In cooling systems, corrosion causes two basic problems. The first and most obvious is the failure of equipment with the resultant cost of replacement and plant downtime. The second is decreased plant efficiency to loss of heat transfer, the result of heat exchanger fouling caused by the accumulation of corrosion products.

Corrosion occurs at the anode, where metal dissolves. Often, this is separated by a physical distance from the cathode, where a reduction reaction takes place. An electrical potential difference exists between these sites, and current flows through the solution from the anode to the cathode. This is accompanied by the flow of electrons from the anode to the cathode through the metal (Fig. 8).

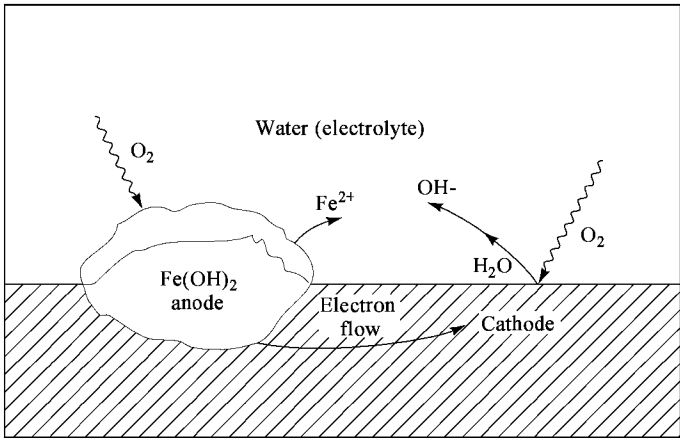
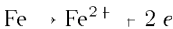


Fig. 8. A corrosion cell.

For steel, the typical anodic oxidation reaction is

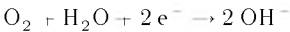


This reaction is accompanied by the following:



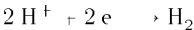
The ferrous hydroxide then combines with oxygen and water to produce ferric hydroxide, Fe(OH)₃, which becomes common iron rust when dehydrated to Fe₂O₃.

The primary cathodic reaction in cooling systems is



The production of hydroxide ions creates a localized high pH at the cathode, approximately 1–2 pH units above bulk water pH. Dissolved oxygen reaches the surface by diffusion, as indicated by the wavy lines in Figure 8. The oxygen reduction reaction controls the rate of corrosion in cooling systems; the rate of oxygen diffusion is usually the limiting factor.

Another important cathodic reaction is



At neutral or higher pH, the concentration of H⁺ ions is too low for this reaction to contribute significantly to the overall corrosion rate. However, as pH decreases, this reaction becomes more important until, at a pH of about 4, it becomes the predominant cathodic reaction.

Types of Corrosion

The formation of anodic and cathodic sites, necessary to produce corrosion, can occur for any of a number of reasons: impurities in the metal, localized stresses, metal grain size or composition differences, discontinuities on the surface, and differences in the local environment (eg, temperature, oxygen, or salt concentration). When these local differences are not large and the anodic and cathodic sites can shift from place to place on the metal surface, corrosion is uniform. With uniform corrosion, fouling is usually a more serious problem than equipment failure.

Localized corrosion, which occurs when the anodic sites remain stationary, is a more serious industrial problem. Forms of localized corrosion include pitting, selective leaching (eg, dezincification), galvanic corrosion, crevice or underdeposit corrosion, intergranular corrosion, stress corrosion cracking, and microbiologically influenced corrosion. Another form of corrosion, which cannot be accurately categorized as either uniform or localized, is erosion corrosion.

Pitting. Pitting is one of the most destructive forms of corrosion and also one of the most difficult to predict in laboratory tests (Fig. 9). Pitting occurs when anodic and cathodic sites become stationary due to large differences in surface conditions. It is generally promoted by low velocity or stagnant conditions (eg, shellside cooling) and by the presence of chloride ions. Once a pit is formed, the solution inside it is isolated from the bulk environment and becomes increasingly corrosive. The high corrosion rate in the pit produces an excess of positively charged metal cations, which attract chloride anions. In addition, hydrolysis produces H⁺ ions. The increase in acidity and concentration within the pit promotes even higher corrosion rates, and the process becomes self-sustaining. Inhibitors can be used to control pitting, but they must be applied correctly.

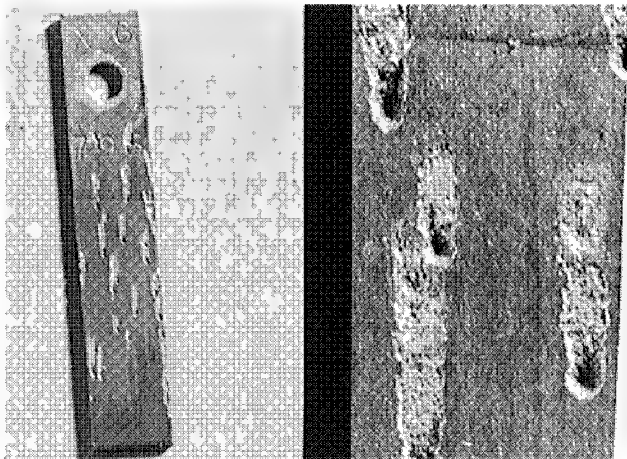


Fig. 9. Pitting corrosion is damaging because it can lead rapidly to equipment failure.

Selective Leaching. Selective leaching is the corrosion of one element of an alloy. The most common example in cooling systems is dezincification, which is the selective removal of zinc from copper–zinc alloys. The conditions that promote the pitting of steel also promote the pitting of brass, which in cooling systems usually occurs by dezincification. Low pH conditions (<6.0) and high free chlorine residuals (<1.0 ppm) are particularly aggressive in producing dezincification. The dezincification resistance varies with the alloy. For example, 70–30 brass is less resistant than Admiralty brass (70–30 brass plus 1^o % tin), which is less resistant than inhibited Admiralty brass (Admiralty brass plus a small amount of arsenic, antimony, or phosphorus).

Galvanic Corrosion. Galvanic corrosion occurs when two dissimilar metals are in contact in a solution. The contact must be good enough to conduct electricity, and both metals must be exposed to the solution. The driving force for galvanic corrosion is the electric potential difference that develops between two metals. This difference increases as the distance between the metals in the galvanic series increases.

Table 4 shows a galvanic series for some commercial metals and alloys. When two metals from the series are in contact in solution, the corrosion rate of the more active (anodic) metal increases and the corrosion rate of the more noble (cathodic) metal decreases.

Table 4. Galvanic Series of Metals and Alloys^a

Corroded End (anodic, or least noble)	
magnesium	Inconel (active)
magnesium alloys	Hastelloy A
zinc	Hastelloy B
Aluminum 2S	brasses
cadmium	copper
Aluminum 17ST	bronzes
steel or iron	copper–nickel alloys
cast iron	titanium
chromium–iron (active)	Monel
Ni-Resist	silver solder
18-8-Cr-Ni-Fe (active)	nickel (passive)
15-8-3'-Cr-Ni-Mo-Fe (active)	Inconel (passive)
Hastelloy C	chromium–iron (passive)
lead–tin solders	18-8 Cr-Ni-Fe (passive)
lead	18-8-3 Cr-Ni-Mo-Fe (passive)
tin	silver
nickel (active)	graphite
Protected End (cathodic, or most noble)	

^a Courtesy of International Nickel Company, Inc.

Galvanic corrosion can be controlled by the use of sacrificial anodes. This is a common method of controlling corrosion in heat exchangers with Admiralty tube bundles and carbon steel tube sheets and channel heads. The anodes are bolted directly to the steel and protect a limited area around the anode. Proper placement of sacrificial anodes is a precise science.

The most serious form of galvanic corrosion occurs in cooling systems that contain both copper and steel alloys. It results when dissolved copper plates onto a steel surface and induces rapid galvanic attack of the steel. The amount of dissolved copper required to produce this effect is small and the increased corrosion is difficult to inhibit once it occurs. A copper corrosion inhibitor is needed to prevent copper dissolution.

Crevice Corrosion. Crevice corrosion is intense localized corrosion that occurs within a crevice or any area that is shielded from the bulk environment. Solutions within a crevice are similar to solutions within a pit in that they are highly concentrated and acidic. Because the mechanisms of corrosion in the two processes are virtually identical, conditions that promote pitting also promote crevice corrosion. Alloys that depend on oxide films for protection (eg, stainless steel and aluminum) are highly susceptible to crevice attack because the films are destroyed by high chloride ion concentrations and low pH. This is also true of protective films induced by anodic inhibitors.

The best way to prevent crevice corrosion is to prevent crevices. From a cooling water standpoint, this requires the prevention of deposits on the metal surface. Deposits may be formed by suspended solids (eg, silt, silica) or by precipitating species, such as calcium salts.

Intergranular Corrosion. Intergranular corrosion is localized attack that occurs at metal grain boundaries. It is most prevalent in stainless steels that have been improperly heat treated. In these metals, the grain boundary area is depleted in chromium and, therefore, is less resistant to corrosion. Intergranular corrosion also occurs in certain high strength aluminum alloys. In general, it is not of significance in cooling systems.

Stress Corrosion Cracking. Stress corrosion cracking (SCC) is the brittle failure of a metal by cracking under tensile stress in a corrosive environment. Failures tend to be transgranular, although intergranular failures have been noted. Commonly used cooling system alloys that may crack due to stress include austenitic stainless steels (300 series) and brasses. The susceptibility of stainless steels to SCC increases significantly as the temperature is increased. Most laboratory stainless steel SCC testing is done at about 150°C, because it is difficult to promote cracking at temperatures below 93°C. For this reason, SCC of stainless steels has not been widely observed in cooling systems.

Chloride is the main contributor to SCC of stainless steels. High chloride concentrations, resulting from high chloride levels in the makeup water and or high cycles of concentration, will increase susceptibility. Although low water temperatures generally preclude cracking, SCC of stainless steels can occur in cooling systems.

For brasses, the ammonium ion is the principal cause of SCC. Few service failures have been reported when ammonia is not present.

The most likely places for SCC to be initiated are crevices or areas where the flow of water is restricted. This is due to the buildup of corrosive concentrations in these areas. For example, chloride can concentrate from 100 ppm in the bulk water to as high as 10,000 ppm (1%) in a crevice. Deposits are initiating sites because of crevices formed beneath them. The low water velocities in shell-side cooling are also detrimental.

The most effective way to prevent SCC in both stainless steel and brass systems is to keep the system clean and free of deposits. An effective deposit control treatment is imperative. A good corrosion inhibitor is also beneficial. Chromate and phosphate have each been used successfully to prevent the SCC of stainless steel in chloride solutions.

Microbiologically Influenced Corrosion (MIC). Microorganisms in cooling water form "biofilms" on cooling system surfaces. Biofilms consist of populations of sessile organisms and their hydrated polymeric secretions. Numerous types of organisms may exist in any particular biofilm, ranging from strictly aerobic bacteria at the water interface to anaerobic bacteria such as sulfate-reducing bacteria (SRB) at the oxygen-depleted metal surface. The presence of a biofilm can contribute to corrosion in three ways: physical deposition, production of corrosive by-products, and depolarization of the corrosion cell caused by chemical reactions.

As discussed above, deposits can cause accelerated localized corrosion by creating differential aeration cells. This same phenomenon occurs with a biofilm. The nonuniform nature of biofilm formation creates an inherent differential, which is enhanced by the oxygen consumption of organisms in the biofilm.

Many of the by-products of microbial metabolism, including organic acids and hydrogen sulfide, are corrosive. These materials can concentrate in the biofilm, causing accelerated metal attack. Corrosion tends to be self-limiting due to the buildup of corrosion reaction products. However, microbes can absorb some of these materials in their metabolism, thereby removing them from the anodic or cathodic site. The removal of reaction products, termed *depolarization*, stimulates further corrosion. Figure 10 shows a typical result of microbial corrosion. The surface exhibits scattered areas of localized corrosion, unrelated to flow pattern. The corrosion appears to spread in a somewhat circular pattern from the site of initial colonization.

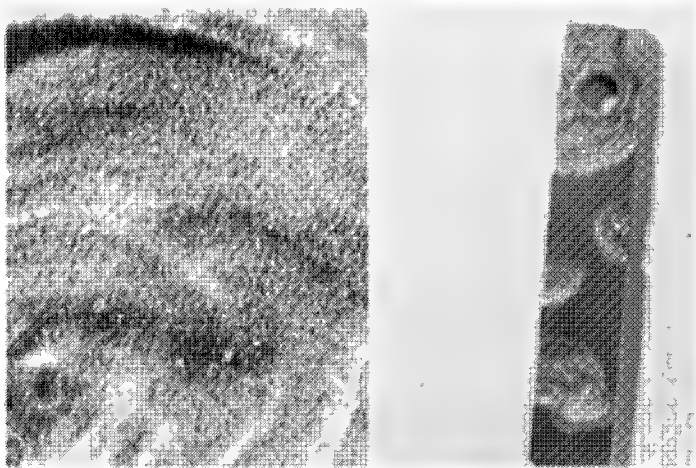


Fig. 10. Microbiologically influenced corrosion (MIC).

Erosion Corrosion. Erosion corrosion is the increase in the rate of metal deterioration from abrasive effects. It can be identified by grooves and rounded holes, which usually are smooth and have a directional pattern. Erosion corrosion is increased by high water velocities and suspended solids. It is often localized at areas where water changes direction. Cavitation (damage due to the formation and collapse of bubbles in high velocity turbines, propellers, etc) is a form of erosion corrosion. Its appearance is similar to closely spaced pits, although the surface is usually rough.

Control of Cooling System Corrosion

Corrosion control requires a change in either the metal or the environment. The first approach, changing the metal, is expensive. Also, highly alloyed materials, which are resistant to general corrosion, are more prone to failure by localized corrosion mechanisms such as stress corrosion cracking.

The second approach, changing the environment, is a widely used, practical method of preventing corrosion. In aqueous systems, there are three ways to effect a change in environment to inhibit corrosion: (1) form a protective film of calcium carbonate on the metal surface using the natural calcium and alkalinity in the water, (2) remove the corrosive oxygen from the water, either by mechanical or chemical deaeration, and (3) add corrosion inhibitors.

Calcium Carbonate Protective Scale. The Langelier saturation index (LSI) is a useful tool for predicting the tendency of a water to deposit or dissolve calcium carbonate. Work published in 1936 deals with the conditions at which a water is in equilibrium with calcium carbonate. An equation developed by Langelier makes it possible to predict the tendency of calcium carbonate either to precipitate or to dissolve under varying conditions. The equation expresses the relationship of pH, calcium, total alkalinity, dissolved solids, and temperature as they relate to the solubility of calcium carbonate in waters with a pH of 6.5–9.5:

$$pH_s = (pK_2 - pK_s) + pCa^{2+} + pAlk$$

where pH_s is the pH at which water with a given calcium content and alkalinity is in equilibrium with calcium carbonate, K_2 is the second dissociation constant for carbonic acid, and K_s is the solubility product constant for calcium carbonate. These terms are functions of temperature and total mineral content. Their values for any given condition can be computed from known thermodynamic constants. Both the calcium ion and the alkalinity terms are the negative logarithms of their respective concentrations. The calcium content is molar, while the alkalinity is an equivalent concentration (ie, the titratable equivalent of base per liter). The calculation of the pH_s has been simplified by the preparation of various nomographs. A typical one is shown in Figure 11.

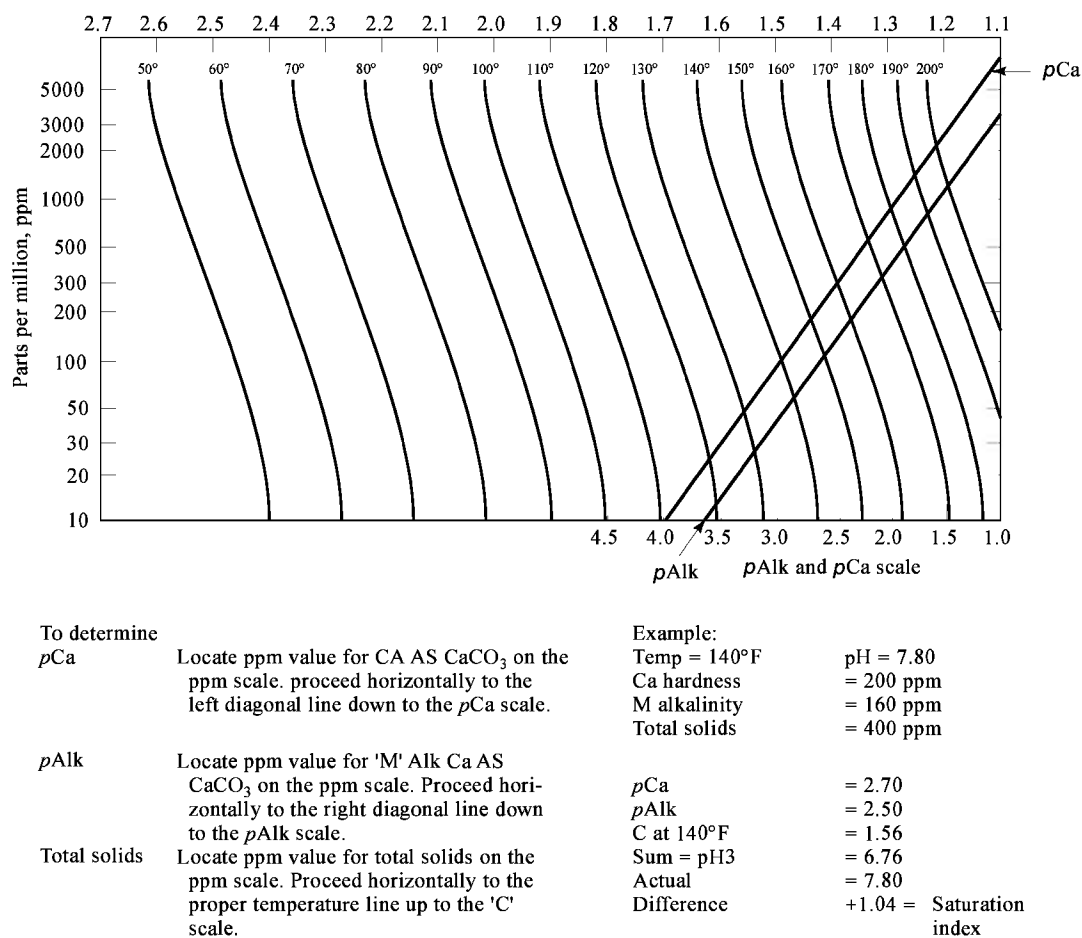


Fig. 11. A typical Langelier Saturation Index chart.

The difference between the actual pH (pH_a) of a sample of water and the pH_i ($pH_a - p H_i$) is called the Langelier saturation index. This index is a qualitative indication of the tendency of calcium carbonate to deposit or dissolve. If the LSI is positive, calcium carbonate tends to deposit. If it is zero, the water is at equilibrium.

The LSI measures only the directional tendency or driving force for calcium carbonate to precipitate or dissolve. It cannot be used as a quantitative measure. Two different waters, one of low hardness (corrosive) and the other of high hardness (scale-forming), can have the same saturation index.

The stability index developed by Ryzner makes it possible to distinguish between two such waters. This index is based on a study of actual operating results with waters having various saturation indexes.

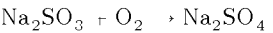
Stability index = 2(pH_s) - pH_a

Where waters have a stability index of 6.0 or less, scaling increases and the tendency to corrode decreases. Where the stability index exceeds 7.0, scaling may not occur at all. As the stability index rises above 7.5 or 8.0, the probability of corrosion increases. Use of the LSI together with the Stability Index contributes to more accurate prediction of the scaling or corrosive tendencies of a water.

A uniform coating of calcium carbonate deposited on the metal surfaces physically segregates the metal from the corrosive environment. To develop the positive LSI required to deposit calcium carbonate, it is usually necessary to adjust the pH or calcium content of the water. Soda ash, caustic soda, or lime (calcium hydroxide) may be used for this adjustment. Lime is usually the most economical alkali because it raises the calcium content as well as the alkalinity.

Theoretically, controlled deposition of calcium carbonate scale can provide a film thick enough to protect, yet thin enough to allow adequate heat transfer. However, low temperature areas do not permit the development of sufficient scale for corrosion protection, and excessive scale forms in high temperature areas and interferes with heat transfer. Therefore, this approach is not used for industrial cooling systems. Controlled calcium carbonate deposition has been used successfully in some waterworks distribution systems where substantial temperature increases are not encountered.

Mechanical and Chemical Deaeration. The corrosive qualities of water can be reduced by deaeration. Vacuum deaeration has been used successfully in once-through cooling systems. When all oxygen is not removed, catalyzed sodium sulfite can be used to remove the remaining oxygen. The sulfite reaction with dissolved oxygen is



The use of catalyzed sodium sulfite for chemical deaeration requires eight parts of catalyzed sodium sulfite for each part of dissolved oxygen. In certain systems where vacuum deaeration is already used, the application of catalyzed sodium sulfite may be economically justified for removal of the remaining oxygen. The use of sodium sulfite may also be applicable to some closed loop cooling systems. In open recirculating cooling systems, continual replenishment of oxygen as the water passes over the cooling tower makes deaeration impractical.

Corrosion Inhibitors. A corrosion inhibitor is any substance that effectively decreases the corrosion rate when added to an environment. An inhibitor can be identified most accurately in relation to its function: removal of the corrosive substance, passivation, precipitation, or adsorption.

- Deaeration (mechanical or chemical) removes the corrosive substance—oxygen.
- Passivating (anodic) inhibitors form a protective oxide film on the metal surface; they are the best inhibitors because they can be used in economical concentrations and their protective films are tenacious and tend to be rapidly repaired if damaged.
- Precipitating (cathodic) inhibitors are simply chemicals that form insoluble precipitates that can coat and protect the surface; precipitated films are not as tenacious as passive films and take longer to repair after a system upset.
- Adsorption inhibitors have polar properties that cause them to be adsorbed on the surface of the metal; they are usually organic materials.

Passivation Inhibitors. Examples of passivators (anodic inhibitors) include chromate, nitrite, molybdate, and orthophosphate. All are oxidizers and promote passivation by increasing the electrical potential of the iron. Chromate and nitrite do not require oxygen, and thus, can be the most effective. Chromate is an excellent aqueous corrosion inhibitor, particularly from a cost perspective. However, owing to health and environmental concerns, use of chromate has decreased significantly and will probably be outlawed soon. Nitrite is also an effective inhibitor, but in open systems it tends to be oxidized to nitrate.

Both molybdate and orthophosphate are excellent passivators in the presence of oxygen. Molybdate can be an effective inhibitor, especially when combined with other chemicals. Orthophosphate is not really an oxidizer *per se*, but becomes one in the presence of oxygen. If iron is put into a phosphate solution without oxygen present, the corrosion potential remains active and the corrosion rate is not reduced. However, if oxygen is present, the corrosion potential increases in the noble direction and the corrosion rate decreases significantly.

A negative attribute of orthophosphate is its tendency to precipitate with calcium hardness found in natural waters. In recent years, deposit control agents that prevent this deposition have been developed. Owing to its relatively low cost, orthophosphate is widely used as an industrial corrosion inhibitor.

Precipitating Inhibitors. As discussed earlier, the localized pH at the cathode of the corrosion cell is elevated due to the generation of hydroxide ions. Precipitating inhibitors form complexes that are insoluble at this high pH (1–2 pH units above bulk water), but whose deposition can be controlled at the bulk water pH (typically 7–9 pH). A good example is zinc, which can precipitate as hydroxide, carbonate, or phosphate. Calcium carbonate and calcium orthophosphate are also precipitating inhibitors. Orthophosphate thus exhibits a dual mechanism, acting as both an anodic passivator and a cathodic precipitator.

Copper Corrosion Inhibitors. The most effective corrosion inhibitors for copper and its alloys are the aromatic triazoles, such as benzotriazole (BZT) and tolyltriazole (TTA). These compounds bond directly with cuprous oxide (Cu_2O) at the metal surface, forming a "chemisorbed" film. The plane of the triazole lies parallel to the metal surface, thus each molecule covers a relatively large surface area. The exact mechanism of inhibition is unknown. Various studies indicate anodic inhibition, cathodic inhibition, or a combination of the two. Other studies indicate the formation of an insulating layer between the water surface and the metal surface. A recent study supports the idea of an electronic stabilization mechanism. The protective cuprous oxide layer is prevented from oxidizing to the nonprotective cupric oxide. This is an anodic mechanism. However, the triazole film exhibits some cathodic properties as well.

In addition to bonding with the metal surface, triazoles bond with copper ions in solution. Thus dissolved copper represents a "demand" for triazole, which must be satisfied before surface filming can occur. Although the surface demand for triazole filming is generally negligible, copper corrosion products can consume a considerable amount of treatment chemical. Excessive chlorination will deactivate the triazoles and significantly increase copper corrosion rates. Due to all of these factors, treatment with triazoles is a complex process.

Adsorption Inhibitors. Adsorption inhibitors must have polar properties to be adsorbed and block the surface against further adsorption. Typically, they are organic compounds containing nitrogen groups, such as amines, and organic compounds containing sulfur or hydroxyl groups. The size, orientation, shape, and electrical charge distribution of the molecules are all important factors. Often, these molecules are surfactants and have dual functionality. They contain a hydrophilic group, which adsorbs onto the metal surface, and an opposing hydrophobic group, which prevents further wetting of the metal.

Glycine derivatives and aliphatic sulfonates are examples of compounds that can function in this way. The use of these inhibitors in cooling systems is usually limited by their biodegradability and their toxicity toward fish. In addition, they can form thick, oily surface films, that may severely retard heat transfer.

Silicates. For many years, silicates have been used to inhibit aqueous corrosion, particularly in potable water systems. Probably due to the complexity of silicate chemistry, their mechanism of inhibition has not yet been firmly established. They are nonoxidizing and require oxygen to inhibit corrosion, so they are not passivators in the classical sense. Yet they do not form visible precipitates on the metal surface. They appear to inhibit by an adsorption mechanism. It is thought that silica and iron corrosion products interact. However, recent work indicates that this interaction may not be necessary. Silicates are slow-acting inhibitors; in some cases, 2 or 3 weeks may be required to establish protection fully. It is believed that the polysilicate ions or colloidal silica are the active species and these are formed slowly from monosilicic acid, which is the predominant species in water at the pH levels maintained in cooling systems.

Cooling System Deposits

Deposit accumulations in cooling water systems reduce the efficiency of heat transfer and the carrying capacity of the water distribution system. In addition, the deposits cause oxygen differential cells to form. These cells accelerate corrosion and lead to process equipment failure. Deposits range from thin, tightly adherent films to thick, gelatinous masses, depending on the depositing species and the mechanism responsible for deposition.

Deposit formation is influenced strongly by system parameters, such as water and skin temperatures, water velocity, residence time, and system metallurgy. The most severe deposition is encountered in process equipment operating with high surface temperatures and or low water velocities. With the introduction of high efficiency film fill deposit accumulation in the cooling tower packing has become an area of concern. Deposits are broadly categorized as scale or foulants.

Scale. Scale deposits are formed by precipitation and crystal growth at a surface in contact with water. Precipitation occurs when solubilities are exceeded either in the bulk water or at the surface. The most common scale-forming salts that deposit on heat transfer surfaces are those that exhibit retrograde solubility with temperature.

Although they may be completely soluble in the lower temperature bulk water, these compounds (eg, calcium carbonate, calcium phosphate, and magnesium silicate) supersaturate in the higher temperature water adjacent to the heat-transfer surface and precipitate on the surface.

Scaling is not always related to temperature. Calcium carbonate and calcium sulfate scaling occur on unheated surfaces when their solubilities are exceeded in the bulk water. Metallic surfaces are ideal sites for crystal nucleation because of their rough surfaces and the low velocities adjacent to the surface. Corrosion cells on the metal surface produce areas of high pH, which promote the precipitation of many cooling water salts. Once formed, scale deposits initiate additional nucleation, and crystal growth proceeds at an accelerated rate.

Scale control can be achieved through operation of the cooling system at subsaturated conditions or through the use of chemical additives. The most direct method of inhibiting formation of scale deposits is operation at subsaturation conditions, where scale-forming salts are soluble. For some salts, it is sufficient to operate at low cycles of concentration and or control pH. However, in most cases, high blowdown rates and low pH are required so that solubilities are not exceeded at the heat transfer surface. In addition, it is necessary to maintain precise control of pH and concentration cycles. Minor variations in water chemistry or heat load can result in scaling (Fig. 12).

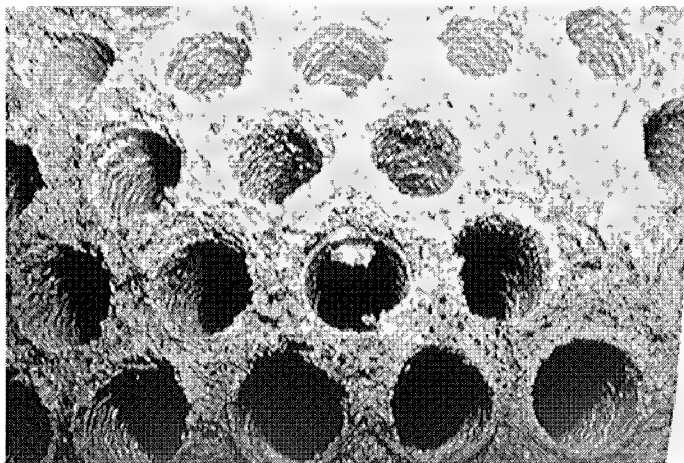


Fig. 12. Calcium carbonate scaling of a surface condenser due to poor pH control.

Threshold Inhibitors. Deposit control agents that inhibit precipitation at dosages far below the stoichiometric level required for sequestration or chelation are called *threshold inhibitors*. These materials affect the kinetics of the nucleation and crystal growth of scale-forming salts and permit supersaturation without scale formation. Threshold inhibitors function by an adsorption mechanism. As ion clusters in solution become oriented, metastable microcrystallites (highly oriented ion clusters) are formed. At the initial stage of precipitation, the microcrystallite can either continue to grow (forming a larger crystal with a well defined lattice) or dissolve. Threshold inhibitors prevent precipitation by adsorbing on the newly emerging crystal, blocking active growth sites. This inhibits further growth and favors the dissolution reaction. The precipitate dissolves and releases the inhibitor, which is then free to repeat the process.

Threshold inhibitors delay or retard the rate of precipitation. Crystals eventually form, depending on the degree of supersaturation and system retention time. After stable crystals appear, their continued growth is retarded by adsorption of inhibitor. The inhibitor blocks much of the crystal surface, causing distortions in the crystal lattice as growth continues. The distortions (defects in the crystal lattice) create internal stresses, making the crystal fragile. Tightly adherent scale deposits do not form, because crystals that form on surfaces in contact with flowing water cannot withstand the mechanical force exerted by the water. The adsorbed inhibitor also disperses particles, by virtue of its electrostatic charge, and prevents the formation of strongly bound agglomerates.

The most commonly used scale inhibitors are low molecular weight acrylate polymers and organophosphorus compounds (phosphonates). Both classes of materials function as threshold inhibitors; however, the polymeric materials are more effective dispersants. Selection of a scale control agent depends on the precipitating species and its degree of supersaturation. The most effective scale control programs use both a precipitation inhibitor and a dispersant. In some cases this can be achieved with a single component (eg, polymers used to inhibit calcium phosphate at near neutral pH).

Fouling. Fouling occurs when insoluble particulates suspended in recirculating water form deposits on a surface. Fouling mechanisms are dominated by particle–particle interactions that lead to the formation of agglomerates. At low water velocities, particle settling occurs under the influence of gravity. Parameters that affect the rate of settling are particle size, relative liquid and particle densities, and liquid viscosity. The relationships of these variables are expressed by Stokes' law. The most important factor affecting the settling rate is the size of the particle. Because of this, the control of fouling by preventing agglomeration is one of the most fundamental aspects of deposition control.

Foulants enter a cooling system with makeup water, airborne contamination, process leaks, and corrosion. Most potential foulants enter with makeup water as particulate matter, such as clay, silt, and iron oxides. Insoluble aluminum and iron hydroxides enter a system from makeup water pretreatment operations. Some well waters contain high levels of soluble ferrous iron that is later oxidized to ferric iron by dissolved oxygen in the recirculating cooling water. Because it is insoluble, the ferric iron precipitates. The steel corrosion process is also a source of ferrous iron and, consequently, contributes to fouling.

Both iron and aluminum are particularly troublesome because of their ability to act as coagulants. Also, their soluble and insoluble hydroxide forms can each cause precipitation of some water treatment chemicals, such as orthophosphate. Airborne contaminants usually consist of clay and dirt particles but can include gases such as hydrogen sulfide, which forms insoluble precipitates with many metal ions. Process leaks introduce a variety of contaminants that accelerate deposition and corrosion.

Foulants, such as river water silt, enter the system as finely dispersed particles, which can be as small as 1–100 nm. The particles carry an electrostatic charge, which causes similarly charged particles to repel each other, favoring their dispersion. The net charge a particle carries depends on the composition of the water. Cycling of cooling water increases the concentration of counter-charged ions capable of being electrostatically attracted to and adsorbed onto a charged particle. As counterions adsorb, the net charge of the particle decreases. Particles begin to agglomerate and grow in size as their repulsive forces are diminished.

Settling occurs when the energy imparted by fluid velocity can no longer suspend the particle, due to agglomeration and growth. After particles have settled, the nature of the deposit depends on the strength of the attractive forces between the particles themselves (agglomerate strength) and between the particles and the surface they contact. If attractive forces between particles are strong and the particles are not highly hydrated, deposits are dense and well structured; if the forces are weak, the deposits are soft and pliable. Deposition continues as long as the shear strength of the deposit exceeds the shear stress of the flowing water.

Removal of Particulate Matter. The amount of particulate entering a cooling system with the makeup water can be reduced by filtration and/or sedimentation processes. Particulate removal can also be accomplished by filtration of recirculating cooling water. These methods do not remove all of the suspended matter from the cooling water. The level of fouling experienced is influenced by the effectiveness of the particular removal scheme employed, the water velocities in the process equipment, and the cycles of concentration maintained in the cooling tower.

High Water Velocities. The ability of high water velocities to minimize fouling depends on the nature of the foulant. Clay and silt deposits are more effectively removed by high water velocities than aluminum and iron deposits, which are more tacky and form interlocking networks with other precipitates. Operation at high water velocities is not always a viable solution to clay and silt deposition because of design limitations, economic considerations, and the potential for erosion corrosion.

Dispersants. Dispersants are materials that suspend particulate matter by adsorbing onto the surface of particles and imparting a high charge. Electrostatic repulsion between like-charged particles prevents agglomeration, which reduces particle growth. The presence of a dispersant at the surface of a particle also inhibits the bridging of particles by precipitates that form in the bulk water. The adsorption of the dispersant makes particles more hydrophilic and less likely to adhere to surfaces. Thus dispersants affect both particle-to-particle and particle-to-surface interactions.

The most effective and widely used dispersants are low molecular weight anionic polymers. Dispersion technology has advanced to the point at which polymers are designed for specific classes of foulants or for a broad spectrum of materials. Acrylate-based polymers are widely used as dispersants. They have advanced from simple homopolymers of acrylic acid to more advanced copolymers and terpolymers. The performance characteristics of the acrylate polymers are a function of their molecular weight and structure, along with the types of monomeric units incorporated into the polymer backbone.

Surfactants. Surface-active or wetting agents are used to prevent fouling by insoluble hydrocarbons. They function by emulsifying the hydrocarbon through the formation of microdroplets containing the surfactant. The hydrophobic (water hating) portion of the surfactant is dissolved within the oil drop, while the hydrophilic (water loving) portion is at the surface of the droplet. The electrostatic charge imparted by hydrophilic groups causes the droplets to repel each other, preventing coalescence. Through a similar process, surfactants also assist in the removal of hydrocarbon-containing deposits.

Cooling water systems, particularly open recirculating systems, provide a favorable environment for the growth of microorganisms. Microbial growth on wetted surfaces leads to the formation of biofilms. If uncontrolled, such films cause fouling, which can adversely affect equipment performance, promote metal corrosion, and accelerate wood deterioration. These problems can be controlled through proper biomonitoring and application of appropriate cooling water antimicrobials.

Biofouling

Microbiological fouling in cooling systems is the result of abundant growth of algae, fungi, and bacteria on surfaces. Once-through and open or closed recirculating water systems may support microbial growth, but fouling problems usually develop more quickly and are more extensive in open recirculating systems. Once-through cooling water streams generally contain relatively low levels of the nutrients essential for microbial growth, so growth is relatively slow. Open recirculating systems scrub microbes from the air and, through evaporation, concentrate nutrients present in makeup water. As a result, microbe growth is more rapid. Process leaks may contribute further to the nutrient load of the cooling water. Reuse of waste water for cooling adds nutrients and also contributes large amounts of microbes to the cooling system.

In addition to the availability of organic and inorganic nutrients, factors such as temperature, normal pH control range, and continuous aeration of the cooling water contribute to an environment that is ideal for microbial growth. Sunlight necessary for growth of algae may also be present. As a result,

large, varied microbial populations may develop.

The outcome of uncontrolled microbial growth on surfaces is "slime" formation. Slimes typically are aggregates of biological and nonbiological materials. The biological component, known as the biofilm, consists of microbial cells and their by-products. The predominant by-product is extracellular polymeric substance (EPS), a mixture of hydrated polymers. These polymers form a gel-like network around the cells and appear to aid attachment to surfaces. The nonbiological components can be organic or inorganic debris from many sources that have become adsorbed to or embedded in the biofilm polymer.

Slimes can form throughout once-through and recirculating systems and may be seen or felt where accessible. In nonexposed areas, slimes can be manifested by decreased heat transfer efficiency or reduced water flow. Wood-destroying organisms may penetrate the timbers of the cooling tower, digesting the wood and causing collapse of the structure. Microbial activity under deposits or within slimes can accelerate corrosion rates and even perforate heat exchanger surfaces.

Microorganisms. The microorganisms that form slime deposits in cooling water systems are common soil, aquatic, and airborne microbes. These microbes may enter the system with makeup water, either in low numbers from fresh water sources or in high numbers when the makeup is waste water. Significant amounts may also be scrubbed from the air as it is drawn through the cooling tower. Process leaks may contribute microorganisms as well.

Bacteria. A wide variety of bacteria can colonize cooling systems. Spherical, rod-shaped, spiral, and filamentous forms are common. Some produce spores to survive adverse environmental conditions such as dry periods or high temperatures. Both aerobic bacteria (which thrive in oxygenated waters) and anaerobic bacteria (which are inhibited or killed by oxygen) can be found in cooling systems.

Fungi. Two forms of fungi commonly encountered are molds (filamentous forms) and yeasts (unicellular forms). Molds can be quite troublesome, causing white rot or brown rot of the cooling tower wood, depending on whether they are cellulolytic (attack cellulose) or lignin degrading. Yeasts are also cellulolytic. They can produce slime in abundant amounts and preferentially colonize wood surfaces.

Algae. Algae are photosynthetic organisms. Green and blue-green algae are common in cooling systems (blue-green algae are now classified with the bacteria and are called cyanobacteria). Various types of algae can be responsible for green growths that block screens and distribution decks. Severe algae fouling can ultimately lead to unbalanced water flow and reduced cooling tower efficiency. Diatoms (algae enclosed by a siliceous cell wall) may also be present but generally do not play a significant role in cooling system problems.

Biofilms. Microbiologists recognize two different populations of microorganisms. Free-floating (planktonic) populations are found in the bulk water. Attached (sessile) populations colonize surfaces. The same kinds of microorganisms can be found in either population, but the sessile population is responsible for biofouling.

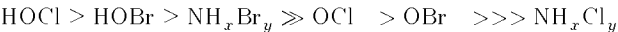
When fouling occurs, even mechanical cleaning does not remove all traces of the biofilm. Previously fouled and cleaned surfaces are more rapidly colonized than new surfaces. Residual biofilm materials promote colonization and reduce the lag time before significant fouling reappears.

Biofilms on heat exchange surfaces act as insulating barriers. Heat exchanger performance begins to deteriorate as soon as biofilm thickness exceeds that of the laminar flow region. Microbes and hydrated biopolymers contain large amounts of water, and biofilms can be more than 90% water by weight. As a result, biofilms have thermal conductivities close to that of water, and in terms of heat transfer efficiency, a biofilm is the equivalent of a layer of stagnant water along the heat exchange surface.

Biofilms can promote corrosion of fouled metal surfaces in a variety of ways. This is referred to as microbially influenced corrosion. Microbes act as biological catalysts promoting conventional corrosion mechanisms: the simple, passive presence of the biological deposit prevents corrosion inhibitors from reaching and passivating the fouled surface; microbial reactions can accelerate ongoing corrosion reactions; and microbial by-products can be directly aggressive to the metal.

Oxidizing Antimicrobials. The oxidizing antimicrobials commonly used in industrial cooling systems are the halogens, chlorine and bromine, in liquid and gaseous form; organic halogen donors; chlorine dioxide; and to a limited extent, ozone. Oxidizing antimicrobials oxidize or accept electrons from other chemical compounds. Their mode of antimicrobial activity can be direct chemical degradation of cellular material or deactivation of critical enzyme systems within the bacterial cell. An important aspect of antimicrobial efficiency is the ability of the oxidizing agent to penetrate the cell wall and disrupt metabolic pathways.

The relative microbiological control ability of typical halogens is as follows:



Cooling water pH affects oxidizing antimicrobial efficacy. The pH determines the relative proportions of hypochlorous acid and hypochlorite ion or, in systems treated with bromine donors, hypobromous acid and hypobromite ion. The acid forms of the halogens are usually more effective antimicrobials than the dissociated forms. Under some conditions, hypochlorous acid is 80 times more effective in controlling bacteria than the hypochlorite ion. Hypochlorous acid predominates below a pH of 7.6. Hypobromous acid predominates below pH 8.7, making bromine donors more effective than chlorine donors in alkaline cooling waters, especially where contact time is limited.

Antimicrobial efficacy is also affected by demand in the cooling water system, specifically demand exerted by ammonia. Chlorine reacts with ammonia to form chloramines, which are not as efficacious as hypochlorous acid or the hypochlorite ion in microbiological control. Bromine reacts with ammonia to form bromamines. Unlike chloramines, bromamines are unstable and reform hypobromous acid.

Most microbes in cooling systems can be controlled by chlorine or bromine treatment if exposed to a sufficient residual for a long enough time. A free chlorine residual of 0.1–0.5 ppm is adequate to control bulk water organisms if the residual can be maintained for a sufficient period of time.

Continuous chlorination of a cooling water system often seems most prudent for microbial slime control. However, it is economically difficult to maintain a continuous free residual in some systems, especially those with process leaks. In some high demand systems it is often impossible to achieve a free residual, and a combined residual must be accepted. In addition, high chlorine feed rates, with or without high residuals, can increase system metal corrosion and tower wood decay. Supplementing with nonoxidizing antimicrobials is preferable to high chlorination rates.

Sodium hypochlorite and calcium hypochlorite are chlorine derivatives formed by the reaction of chlorine with hydroxides. The application of hypochlorite to water systems produces the hypochlorite ion and hypochlorous acid, just as the application of chlorine gas does.

Halogen donors are chemicals that release active chlorine or bromine when dissolved in water. After release, the halogen reaction is similar to that of chlorine or bromide from other sources. Solid halogen donors commonly used in cooling water systems include 1-bromo-3-chloro-5,5-dimethylhydantoin, 1,3-dichloro-5,5-dimethylhydantoin, and sodium dichloroisocyanurate.

Chlorine dioxide, ClO₂, is another chlorine derivative. This unstable, potentially explosive gas must be generated at the point of application. The most common method of generating ClO₂ is through the reaction of chlorine gas with a solution of sodium chlorite.

Ozone. Ozone is an allotropic form of oxygen, O₃. Because it is an unstable gas, it must be generated at the point of use. Ozone is an effective, clean oxidizing agent possessing powerful antibacterial and antiviral properties.

Nonoxidizing Antimicrobials. Nonoxidizing antimicrobials usually control growths by one of two mechanisms. In one, microbes are inhibited or killed as a result of damage to the cell membrane. In the other, microbial death results from damage to the biochemical machinery involved in energy production or energy utilization.

Quaternary ammonium compounds (quats) are cationic surface-active molecules. They damage the cell membranes of bacteria, fungi, and algae. As a result, compounds that are normally prevented from entering the cell are able to penetrate this permeability barrier. Conversely, nutrients and essential intracellular components concentrated within the cell leak out. Growth is hindered, and the cell dies. At low concentrations, quats are biostatic because many organisms can survive in a damaged state for some time. However, at medium to high concentrations, quats can control the organisms.

Many antimicrobials interfere with energy metabolism. Because all microbial activity ultimately depends on the orderly transfer of energy, it can be anticipated that interference with the many energy-yielding or energy-trapping reactions will have serious consequences for the cell. Antimicrobials known to inhibit energy metabolism include organotin, bis(trichloromethyl) sulfone, methylenebis-(thiocyanate) (MBT), β-bromo-β-nitrostyrene (BNS), dodecylguanidine salts, and bromonitropropanediol (BNPD). All of these compounds are effective when applied in sufficient concentrations.

Dodecylguanidine salts also have surfactant properties, which probably contribute to their effectiveness.

Macrofouling Organisms

Fouling caused by large organisms, such as oysters, mussels, clams, and barnacles, is referred to as macrofouling. Typically, organisms are a problem only in large once-through cooling systems or low cycle cooling systems that draw cooling water directly from natural water sources. Water that has been processed by an influent clarification and disinfection system is usually free of the larvae of macrofouling organisms.

Macrofouling has always been a concern in certain regions of the United States, especially in coastal areas. However, in the last 10 years, the incidence of problems in the United States caused by macrofouling has increased dramatically. This is due primarily to the "invasion" of two organisms that were accidentally introduced to this country: the Asiatic clam and the zebra mussel. Both organisms have flourished and represent a significant threat to system reliability. Adding to the problem is the decreased use of chlorine and heavy metal antimicrobials, which permits the infiltration and growth of macrofouling organisms in plant water systems.

Asiatic Clams

Asiatic clams are freshwater mollusks. They probably originated in China or eastern Asia and were introduced into North America and Europe in the past century. They were originally found in warm water but their territory now extends to Minnesota. They have not yet been seen in Canadian rivers or lakes.

Asiatic clams do not attach to surfaces but burrow into sediments in their natural environment. Larvae and juvenile clams easily pass through intake screens (Fig. 13) and settle in low flow areas. Within 6 months to 1 yr, the clams grow to 1.5–2.54 cm in size. When a clam dies, the shell gapes open. Shells of living or dead organisms are carried by water flow and can wedge in condenser or heat exchanger tubes. Once a shell is wedged in a tube, other shells and debris collect and plug the tube further (Fig. 14). The Asiatic clam reaches adulthood in about 1 yr and reproduces in warm months, releasing thousands of larvae into the system.

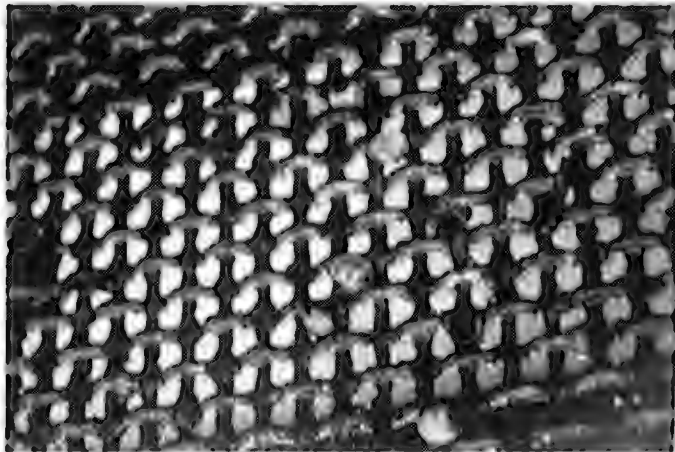


Fig. 13. Juvenile Asiatic clams passing through a water intake screen.



Fig. 14. Surface condenser infested with adult Asiatic clams.

Thermal Treatment (Thermal Backwash). The organisms that cause macrofouling can be killed by heated water. Some systems are designed to allow the heated water from the outlet of the condenser to be recirculated back to the intake. As the water recirculates, it is heated and improves macrofouling control. A 15- to 60-min exposure to water at 40°C or higher has effectively controlled zebra mussels. Thermal treatment is not used extensively because most systems are not designed to recirculate water. Also, when heated water is recirculated the system cooling capacity is greatly diminished.

Oxidizing Antimicrobials. The application of an oxidizing antimicrobial such as chlorine for the control of undesirable organisms is a well-known and long-practiced procedure. Chlorine is toxic to all living organisms from bacteria to humans. However, in the case of hard-shelled creatures, including some mollusks and crustaceans, exposure is not easily accomplished. Some mollusks (eg, oysters, blue mussels, Asiatic clams, and zebra mussels) and crustaceans (eg, barnacles) have sensitive chemoreceptors that detect the presence of oxidizing chemicals such as chlorine (hypochlorite), bromine (hypobromite), ozone, and hydrogen peroxide. When oxidizers are detected at life-threatening levels, the animal withdraws into its shell and closes up tightly to exclude the hostile environment. Animals like oysters and mussels can remain closed for days to weeks if necessary. There is evidence to suggest that, during extended periods of continuous chlorination, the creatures may eventually die from asphyxiation rather than chlorine toxicity.

Even when their shells are closed, the animals continue to sense their environment, and as soon as the oxidant level decreases, they reopen and resume siphoning. Continuous chlorination often fails to eradicate these macrofouling creatures because of interruptions in the feed, which can occur for various reasons, such as chlorine tank changeover or plugging of feedlines. If the interruption lasts long enough (1 h or possibly less), the animals have time to reoxygenate their tissues between the extended periods of chlorination. Any oxidant, such as chlorine, bromine, or ozone, elicits the same response from these creatures. Therefore, only continuous, uninterrupted applications are successful.

Nonoxidizing Antimicrobials. There are several categories of nonoxidizing antimicrobials that have proven to be effective in controlling macrofouling organisms. Quaternary amine compounds and certain surfactants have been applied to infested systems for relatively short intervals (6–48 h). These compounds do not trigger the chemoreceptors of the mollusks. The mollusks continue to filter feed and ingest a lethal dosage of the antimicrobial throughout the exposure period. These compounds produce a latent mortality effect; the mollusks may not die until several hours after the antimicrobial application. Cold water temperatures may extend this latent mortality effect and may also require slightly higher feed rates and longer feed durations due to slower organism metabolism. The advantages of a nonoxidizing antimicrobial program include ease of handling, short application time, and relatively low toxicity to other aquatic organisms. In addition, some of these compounds can be readily detoxified.

BIBLIOGRAPHY

"Water (Industrial)" in *ECT* 1st ed., Vol. 14, pp. 926–946, by E. P. Partridge, Hall Laboratories, Inc.; "Water, Industrial Water Treatment" in *ECT* 2nd ed., Vol. 22, pp. 65–82, by J. K. Rice and D. E. Simon II, Cyrus Wm. Rice Division, NUS Corporation, "Water (Industrial Treatment)" in *ECT* 3rd ed., Vol. 24, pp. by S. Sussman, Olin Water Services.

General References

J. F. Malina and F. G. Pohland, *Design of Anaerobic Processes for the Treatment of Industrial and Municipal Wastes*, Technomic Publishing, Lancaster, Pa., 1992.
C. E. Hamilton, ed., *Manual on Water*, ASTM Special Publication 442A, American Society for Testing and Materials, Philadelphia, Pa., 1978, 472 pp.
Standard Methods for the Examination of Water and Wastewater, 15th ed., American Public Health Association, Washington, D.C.
F. van der Leeden and co-workers, *The Water Encyclopedia*, Lewis Publishers, Chelsea, Mich., 1990.
The Nalco Water Handbook, McGraw-Hill, Inc. Publishers, New York, NY, 1979.
Betz Handbook of Industrial Water Conditioning, Ninth Edition, Trevoise, PA, 1991.
The ASME Handbook on Water Technology for Thermal Power Systems, The American Society of Mechanical Engineers, New York, NY, 1989.
Metals Handbook, Ninth Edition, Volume 13, ASM International, Metals Park, Ohio, 1987.
"A Practical Guide to Avoiding Steam Purity Problems in The Industrial Plant," *CRTD* **35**, (1995).
Consensus on Operating Practices for The Control of Feedwater and Boiler Water Chemistry in Modern Industrial Boilers, The American Society of Mechanical Engineers, New York, NY, 1994.

James Robinson
Betz Dearborn

MUNICIPAL WATER TREATMENT

Sedimentation and Filtration

Most surface waters contain varying amounts of suspended solids, including silt, clay, bacteria, and viruses; and it is necessary to remove these before to distribution to the domestic or industrial consumer. Suspended solids not only affect the acceptability of the water but also interfere with disinfection. The principal treatment processes are sedimentation (qv) and filtration (qv). Sedimentation alone is rarely adequate for the clarification of turbid waters and is of little or no value for the removal of such very fine particles as clay, bacteria, etc. Table 1 shows the effect of particle size on the sedimentation rate of a solid having a specific gravity of 2.65 in water at 20°C.

Table 1. Sedimentation Rate as a Function of Particle Diameter ^a

Equivalent spherical radius	Approximate size	Sedimentation rate (time to settle 30 cm)
10 mm	gravel	0.3 s
1 mm	coarse sand	3 s
100 μ	fine sand	38 s
10 μ	silt	33 min
1 μ	bacteria	55 h
100 nm	colloid	230 h
10 nm	colloid	6.3 yr
1 nm	colloid	63 yr

^a Ref. 1.

In many plants that treat surface waters, there is a presedimentation reservoir ahead of the treatment units. The reservoir allows the larger particles to settle as well as to provide a volume buffer against changes in water quality. Table 1 shows that the only particles that settle within a reasonable period of time are silt, sand, and possibly some larger bacteria. The listed values are for an undisturbed system and the times would be much longer for actual cases where currents tend to upset the settling. Further treatment, which is necessary to produce potable water, may involve only filtration through sand or multiple-medium filters or may require considerable pretreatment, eg, coagulation and flocculation (qv) before filtration.

The use of sand filtration as a method of clarifying water dates to the small sand beds designed for Paisley, Scotland, in 1804 and, later, to the larger sand filters used by the Chelsea Water Company of London in 1828 to filter Thames River water. The first use of sand filters in the United States was in Poughkeepsie, New York, in 1872 (2). These were slow sand filters, ie, the filter rates were low compared to the filter rates of the 1990s. Table 2 compares the properties of slow, rapid, and high rate filters. Slow sand filters have large surface areas because of the low rates that must be used. Higher rate filters were developed because of the very large surface areas that would be required for the filtration of water for larger cities. There are, however, slow sand filters still in use both in the United States and in Europe (4).

Table 2. Properties of Filter Types^a

Property	Type of filter		
	Slow sand	Rapid sand	High rate
design filtration rate, L (sm ²) ^b	0.081–0.162	1.36	2.72–6.83
bed size to filter water 1 m ³ (m ² ·min)	143	1.15	0.39
of filter medium, m ^{2c}			
medium type	ungraded	carefully graded sand with larger sizes at bottom	several types but usually carefully graded sand and carbon
bed depth			
lower layer, cm	30 gravel	30–45 gravel	30–45 sand
top layer, cm	90–110 sand	60–75 sand	45–60 granular carbon
period of filter use be-tween cleanings, d	20–60	1–3	0.5–3
cleaning procedure	removal of surface layer for clean-ing or washing surface	backwash by forcing water up through filter; surface or submerged jets to fluidize bed with water or air	same as rapid sand
water for cleaning, ^o % of filtered water	0.2–0.6	1–6	1–3
depth of filtration	surface only	first few cm	in depth
pretreatment	none	coagulation and sedimentation	coagulation and sedimentation

^a Ref. 3.

^b To convert L (sm²) to gal (sft²), divide by 40.74.

^c To convert m² to ft², multiply by 10.76.

The first patent relating to rapid sand filters was issued in 1884 for the use of ferric chloride as a coagulant ahead of the filters (5). The coagulant duplicates the slimy, gelatinous layer of bacteria, algae, and fungi that built up on the surface of slow sand filters and played an important part in the filtering action. This layer is called the *schmutzdecke* (German, "dirty layer"). The first large, municipal, rapid sand filter was built at Little Falls, New Jersey, in 1909 (2).

Dual-medium filters for municipal and industrial application have been described (6,7). One of the first designs involved an inverted-medium

loading, in which very fine, dense garnet sand was used at the bottom of the filter; this was covered by less dense but coarser silica sand, and then by granulated carbon of even lower density and larger particle size. This type of construction provides larger pore sizes for the initial contact with the water to be filtered and, thereby, allows deeper penetration of smaller particles into the bed and the consequent use of more of the filter medium for filtration. Backwashing returns the filter to the original particle-size distribution because of the density differences of the filter media.

The size of particles removed by such filters is less than the size of the passages. The mechanism of removal includes adsorption (qv) of the impurities at the interface between the media and the water either by specific chemical or van der Waals attractions or by electrostatic interaction when the medium particles have surface charges opposite to those on the impurities to be removed.

Neither rapid sand nor mixed-media filters remove appreciable quantities of colloidal particles without adequate pretreatment. Although it is widely believed that filters are an effective barrier against unsafe water, the effluent may be as colored, as turbid, or as bacteriologically unsafe as the water applied. In contrast, slow sand filters require no pretreatment, as the slow passage through the bed allows the particles to contact and attach to the *schmutzdecke*.

The two steps in the removal of a particle from the liquid phase by the filter medium are the transport of the suspended particle to the surface of the medium and interaction with the surface to form a bond strong enough to withstand the hydraulic stresses imposed on it by the passage of water over the surface. The transport step is influenced by such physical factors as concentration of the suspension, medium particle size, medium particle-size distribution, temperature, flow rate, and flow time. These parameters have been considered in various empirical relationships that help predict filter performance based on physical factors only (8,9). Attention has also been placed on the interaction between the particles and the filter surface. The mechanisms postulated are based on adsorption (qv) or specific chemical interactions (10).

The goal of filtration in the modern municipal treatment plant is a maximum of 0.1 ntu (nephelometric turbidity unit), which ensures a sparkling, clear water (8). Freedom from disease organisms is associated with freedom from turbidity, and complete freedom from taste and odor requires no less than such clarity. The National Interim Primary Drinking Water Regulations (NIPDWR) require that the maximum contaminant level for turbidity at the point of entry into the distribution system be 1.0 ntu unless it can be shown that levels up to 5 ntu do not interfere with disinfection, interfere with the maintenance of a chlorine residual in the distribution system, nor interfere with bacteriological analyses.

Coagulation and Flocculation. The removal of colloidal particles, eg, turbidity, color, and bacteria, requires agglomeration of these particles prior to filtration. Agglomeration can be carried out chemically, that is by interaction of the colloidal particles with materials added as coagulants, or physicochemically by neutralizing the charge on the particles or by interparticle bridging. The term coagulation is applied to the addition of any material that causes agglomeration of the colloids; flocculation refers to the process of gentle agitation which builds floc particles large enough to settle rapidly. A physicochemical mechanism has been proposed for coagulation and is based on charge neutralization by double-layer compression, which allows the particles to approach closely enough for short-range van der Waals forces to cause agglomeration (11). Flocculation is the growth of a three-dimensional structure by interparticle bridging.

The repulsive forces that act to stabilize the suspension are the charge on the particles (like particles have like charges and therefore repel one another) and hydration. The attractive or destabilizing forces are Brownian movement, van der Waals forces, and gravity. As shown in Table 1, the effect of gravity is insignificant in the sedimentation of colloidal particles. Thus, it is possible to destabilize colloids by increasing the attractive forces or by decreasing the repulsive forces.

Among the various factors that result in charges on colloidal particles are ionization of surface groups, eg, acid groups on organic color particles and saltlike bonds on the surface of clay particles that result in ion-exchange reactions with the solution. In general, the colloidal particles in natural waters are negatively charged. Clay, for example, has a net negative charge on the large surface with positive charges on the edges. Color particles and bacteria are also negatively charged. This charge cannot be measured directly, as the particles are surrounded by a sheath of counterions in a relatively fixed layer (Fig. 1). From this fixed layer, the charge is gradually neutralized by the higher concentration of counterions in the diffuse double layer. The measurements that can be made of this charge are limited by the impossibility of obtaining the particle independent on the counterions. The rate at which the particle and some of its counterions move under an impressed potential difference can be measured by various techniques, eg, microelectrophoresis. However, this gives only the charge at the slipping plane or shear plane where the net force between the particle and the bulk solution is small enough so that the particle moves independently. The charge at this plane is referred to as the zeta potential and is shown in Figure 1 as ψ_ζ . The charge on the particle is ψ_0 .

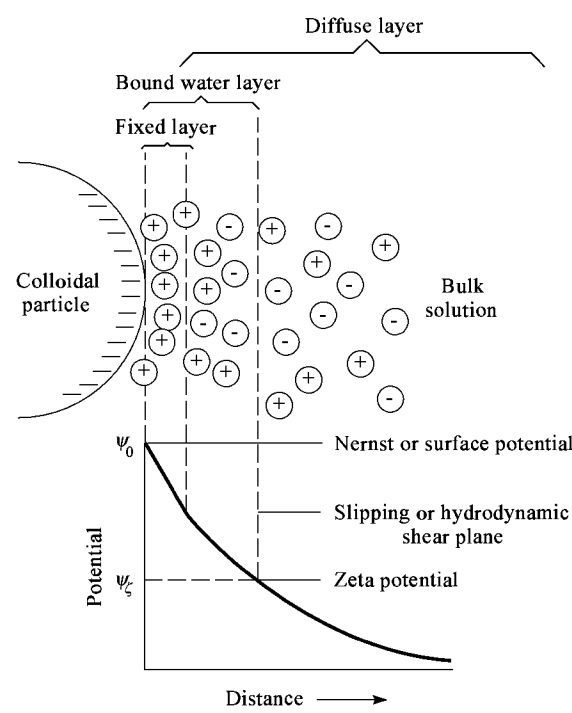


Fig. 1. Double-layer model of colloidal particle.

The other important stabilizing force is hydration, because hydrophilic colloids are involved. The waters of hydration modify the exterior of the particle, so that it approaches the properties of the bulk water. Owing to this, the particles show no tendency to approach or to coalesce. The destabilizing effect of the Brownian movement brings the particles into contact simply because of their random wanderings through the solution. If two particles happen to collide or penetrate each other's repulsive sphere, they may be held together by van der Waals forces. These forces are proportional to the reciprocal of the distance between the particle to the fifth power. This means that attractive forces generally act over a much shorter distance than do repulsive forces.

It can be seen from these two factors, ie, particle charge and van der Waals forces, that the charge must be reduced or the double layer must be compressed to allow the particles to approach each other closely enough so that the van der Waals forces can hold them together. There are two approaches to the accomplishment of this goal: reaction of the charged surface sites with an opposite charge on an insoluble material and neutralization of

the charge by opposite charges concentrated in the fixed layer or the immediate environment. It is difficult to distinguish between these two as they result in the same effect on the particle, insofar as this effect can be measured. Because the particles are negatively charged, they can be coagulated by organic or inorganic cationic polyelectrolytes. A particle that consists of an organic chain that is negatively charged or uncharged may still be adsorbed, but this does not result in reduction of the particle charge.

Many metal ions react with water to produce hydrolysis products that are multiply charged inorganic polymers. These may react specifically with negative sites on the colloidal particles to form relatively strong chemical bonds, or they may be adsorbed at the interface. In either case, the charge on the particle is reduced.

There are two classes of coagulants used in municipal water treatment: inorganic metal salts, containing iron and aluminum, and organic polymers. The metal salts have been used since the early days of water treatment as the primary coagulants, although they were first utilized as pretreatment for filtration. The metal salts hydrolyze when added to water, according to the typical series of reactions shown in Table 3. The predominant equilibrium species for the hydrolysis of aluminum and iron(III) ions over the pH range of interest in water treatment, ie, pH 5–9, are $\text{Al}(\text{OH})_3$ and $\text{Fe}(\text{OH})_3$, as shown in Figures 2 and 3, which express the concentration of the various species as a function of pH. These equilibrium conditions may occur during the coagulation, flocculation, and sedimentation processes in treatment plants. It has been shown that the reactive species may not be those that predominate under equilibrium conditions but rather the less-hydrolyzed species that form in the first few seconds or minutes of the reaction with water (14,15). These species are positively charged and effectively interact with the negative colloidal impurities.

Table 3. Hydrolytic Reactions of Metal Ions

Reaction	p <i>K</i>	Ref.
$\text{Al}^{3+} + \text{H}_2\text{O} \rightleftharpoons \text{AlOH}^{2+} + \text{H}^+$	5.03	12
$2 \text{Al}^{3+} + 4 \text{H}_2\text{O} \rightleftharpoons \text{Al}_2(\text{OH})_4^{2+} + 4 \text{H}^+$	6.27	12
$\text{Al}^{3+} + 3 \text{H}_2\text{O} \rightleftharpoons \text{Al}(\text{OH})_3 + 3 \text{H}^+$	9.1	13
$\text{Al}^{3+} + 4 \text{H}_2\text{O} \rightleftharpoons \text{Al}(\text{OH})_4^- + 4 \text{H}^+$	21.84	13
$\text{Fe}^{3+} + \text{H}_2\text{O} \rightleftharpoons \text{FeOH}^{2+} + \text{H}^+$	3.05	8
$\text{Fe}^{3+} + 2 \text{H}_2\text{O} \rightleftharpoons \text{Fe}(\text{OH})_2^+ + 2 \text{H}^+$	6.31	8
$\text{Fe}^{3+} + 3 \text{H}_2\text{O} \rightleftharpoons \text{Fe}(\text{OH})_3 + 3 \text{H}^+$	0.5 ^a	
$\text{Fe}^{3+} + 4 \text{H}_2\text{O} \rightleftharpoons \text{Fe}(\text{OH})_4^- + 4 \text{H}^+$	18.0	9

^a Calculated from $\text{p}K_w = 14$, $\text{p}K_{\text{sp}} = 42.5$.

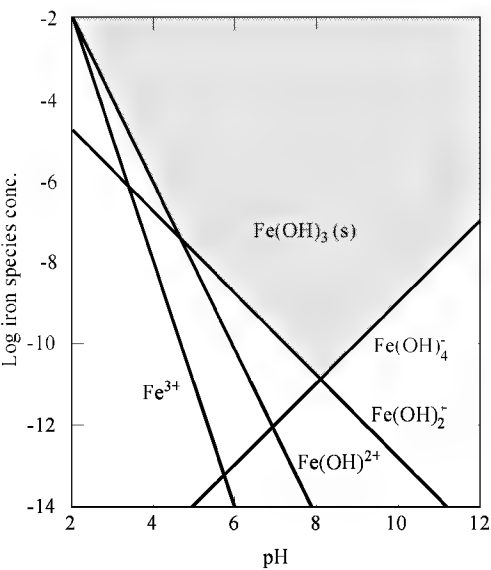


Fig. 2. Equilibrium solubility domain for $\text{Fe}(\text{OH})_3$. The shading defines the area of stability of solid $\text{Fe}(\text{OH})_3$.

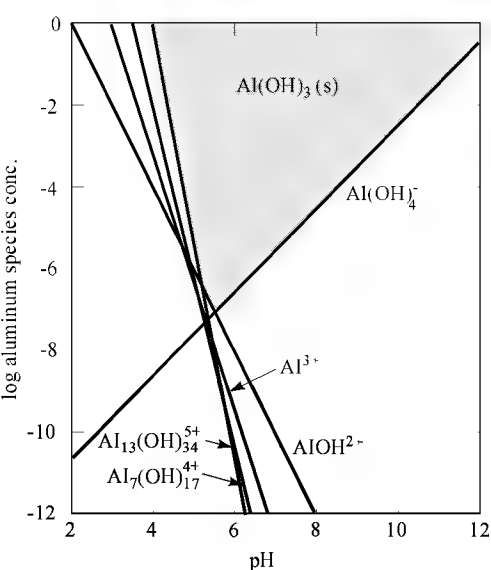


Fig. 3. Equilibrium solubility domain for $\text{Al}(\text{OH})_3$. The shading defines the area of stability of solid $\text{Al}(\text{OH})_3$.

The metal salts reduce the alkalinity in the water; therefore, it may be necessary to add base in the form of lime or soda ash. One part of technical aluminum sulfate ($\text{Al}_2(\text{SO}_4)_3 \cdot 14\text{H}_2\text{O}$) reduces the alkalinity as CaCO_3 by 0.55 parts and one part of technical ferric sulfate, $\text{Fe}_2(\text{SO}_4)_3 \cdot \text{H}_2\text{O}$, by 0.68 parts. The reaction is



The selection of a particular metal salt is based on such factors as local availability, convenience, economics, and effectiveness for the specific treatment problem. Aluminum sulfate (commercial product $\text{Al}_2(\text{SO}_4)_3 \cdot 14\text{H}_2\text{O}$) is available as either a granular solid containing 17 wt % Al_2O_3 or a solution containing 8.3 wt % Al_2O_3 . This is commonly referred to as alum or filter alum. The higher cost of transporting the solution may be more than compensated by the ease in handling and savings in labor and equipment required for feeding and dissolving the solid (16).

Ferric sulfate (commercial product $\text{Fe}_2(\text{SO}_4)_3 \cdot \text{H}_2\text{O}$) contains a minimum of 20 wt % iron(III). It is available only as a solid, which must be dissolved immediately before use. The solution must be kept concentrated to avoid premature hydrolysis and precipitation of $\text{Fe}(\text{OH})_3$. Such concentrated solutions have low pH values and, thus, prevent hydrolysis but are very corrosive. Containers must therefore be coated with or be constructed of corrosion-proof materials.

Ferrous sulfate (commercial product $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$), referred to as copperas, may be the least expensive iron salt, particularly in localities where it may be available as a by-product of the manufacture of titanium dioxide or the pickling of steel. Its high solubility as the hydroxide (ca 4 mg/L as iron) precludes its use as a coagulant; therefore, it must be oxidized to the +3 state. This can be accomplished readily above pH 6 by oxygen or chlorine. Each part of $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ requires 0.03 parts of oxygen or 0.126 parts of chlorine for oxidation to the ferric state.

Sodium aluminate (NaAlO_2) normally contains an excess of sodium hydroxide or soda ash to maintain a sufficiently high pH to prevent $\text{Al}(\text{OH})_3$ precipitation prior to its addition to the water as a coagulant.

The organic coagulants in use may be either natural or synthetic. Among the useful natural polymers are starches, gums, and gelatin. There are four classes of organic polymeric coagulants: cationic, anionic, ampholytic (frequently referred to as polyelectrolytes), and nonionic. The cationic class comprises quaternary ammonium compounds, eg, poly(dimethyldiallylammonium chloride). Poly(sodium methacrylate) is an example of an anionic polyelectrolyte. Poly(lysine–glutamic acid), having both $-\text{NH}_3^+$ and $-\text{COO}^-$ groups, is an example of an ampholytic polyelectrolyte.

Color can be removed effectively and economically with either alum or ferric sulfate at pH values of 5–6 and 3–4, respectively. The reaction is stoichiometric and is a specific reaction of the coagulant with the color to form an insoluble compound (17). The dosage required may be as high as 100–150 mg/L (380–570 mg/gal). Raw-water colors may be as high as 450–500 units on the APHA color scale. The secondary MCL (maximum contaminant level) for color in the finished water is 15 units, although most municipal treatment plants produce water that seldom exceeds 5 units.

The treatment units used for color removal are the same as those used for turbidity removal. However, the pH must be increased prior to filtration so that the metal hydroxides are removed by the filters. At low pH values, metal ions or their soluble complexes readily pass through the filters and form insoluble species in storage tanks and in the distribution system. For iron salts, it is important that the pH be greater than 6 as the oxidation of iron(II) to iron(III) occurs rapidly above this pH in the presence of dissolved oxygen or other strong oxidants (18).

Softening. A water is classified as hard if it contains more than 120 mg of divalent ions per liter (454 mg/L), usually calcium and magnesium, expressed as CaCO_3 . Hard water is problematic as cleaning in it requires a greater amount of soap than is needed in soft water to produce lather. Also, the calcium and magnesium salts that cause hardness have negative solubility coefficients and precipitate with an increase in temperature. A deposit of scale forms on heat-transfer surfaces, leading to localized overheating or clogging of service lines.

The American Water Works Association (AWWA) Water Quality Goals recommend a maximum total hardness of 80 ppm for municipal purposes (19). Municipal softening plants, however, distribute waters containing 70–150 ppm; the final quality is established based on such factors as public demand and economics.

The two principal methods of softening water for municipal purposes are addition of lime or lime-soda and ion exchange. The choice method depends upon such factors as the raw-water quality, the local cost of the softening chemicals, and means of disposing of waste streams.

Lime and Lime–Soda Processes. The first softening plant in the early 1900s used the lime softening process with fill and draw units. Later, continuous-treatment units, which greatly increased the amount of water that could be treated in a facility of given size, were developed. More than 1000 municipalities soften water. Most are in the Midwest and in Florida. However, concern for the adverse effect of soft water on cardiovascular disease (CVD) may limit the number of plants that introduce softening.

The lime or lime–soda process results in the precipitation of calcium as calcium carbonate and magnesium as magnesium hydroxide. The solubilities of these compounds are shown in Figure 4 as functions of pH. When lime is used alone, only the carbonate hardness is reduced. The carbonate hardness is present as calcium or magnesium bicarbonate. The additional use of soda ash can reduce the noncarbonate hardness by providing additional carbonate ion. The reactions involved in the various steps of the process are listed below:

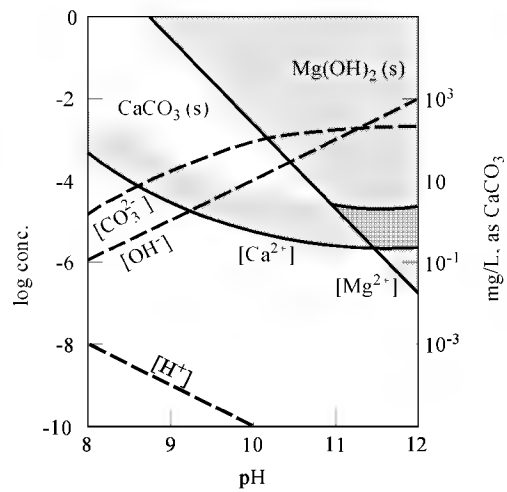


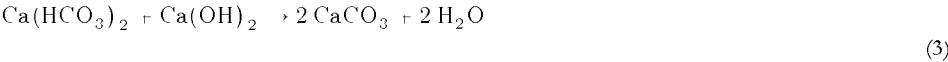
Fig. 4. Equilibrium solubility domains for CaCO_3 and $\text{Mg}(\text{OH})_2$ at a total carbonic species concentration of $5 \times 10^{-3} \text{ M}$. The shaded areas above the Mg^{2+} and Ca^{2+} curves define the areas of stability of solid $\text{Mg}(\text{OH})_2$ and CaCO_3 , respectively.

The reactions of any free CO_2 with the added lime:

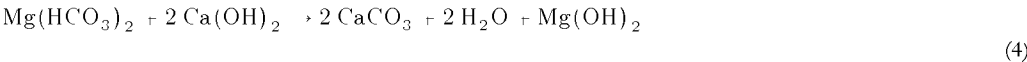


This reaction provides no softening but occurs preferentially, since the CO₂ is the strongest acid present in the system.

The reaction of calcium carbonate hardness with lime, when the calcium can be represented as the bicarbonate at the usual pH values:



The reaction of magnesium carbonate hardness with lime:

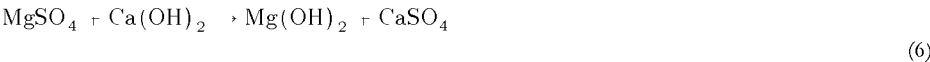


The reaction of calcium noncarbonate hardness with soda ash:



The soda ash provides CO²₃ because available CO²₃ is consumed (eqs. 3–5). The noncarbonate hardness may be represented as sulfate, although any anion except carbonate or bicarbonate could be present.

The reaction of magnesium noncarbonate hardness with lime and with soda ash is a two step reaction since reaction 6 produces a reasonably soluble calcium salt that must react with CO²₃ in order to cause calcium precipitation:



In the presence of soda ash this is followed by equation 5.

From reactions 2–6, it can be seen that the addition of lime always serves three purposes and may serve a fourth. It removes, in order, CO₂, calcium carbonate hardness, and magnesium carbonate hardness (reactions 8, 9, and 10, respectively). Where magnesium noncarbonate hardness must be removed, the lime converts it to calcium noncarbonate hardness (reaction 6). Soda ash, then, removes noncarbonate hardness according to reaction 5.

The needed amounts of lime and soda ash can be calculated from the stoichiometry of the reactions. The effluent quality is a function of the solubilities of calcium carbonate and magnesium hydroxide and of the quantities of softening chemicals added. The acceptable level of total hardness can be decided and usually is 70–120 mg/L (265–454 mg/gal), expressed as CaCO₃. The sum of the solubilities of calcium carbonate and magnesium hydroxide is ca 50–70 mg/L (190–265 mg/gal), depending upon the pH. The sum of the concentrations of the carbonic species HCO[–]₃ + CO²₃, expressed as CaCO₃ or carbonate hardness, can be as low as 20 mg/L (76 mg/gal) but is usually 30–50 mg/L (114–190 mg/gal) because of CaCO₃ supersaturation. It is desirable to reduce the carbonate hardness to the lowest practical value that allows the maximum concentration of noncarbonate hardness to remain. This decreases the cost of treatment, as soda ash costs at least twice as much as lime. Additionally, ca 190% more soda ash than lime is needed to remove the same amount of hardness. It is economically more favorable to minimize the amount of calcium, since twice as much lime is required to remove magnesium as to remove calcium. Unfortunately, all of the hardness left cannot be present as magnesium, since magnesium hardness above 40 ppm causes scaling in hot-water heaters at 16°C, which is the normal temperature setting. For this reason, the magnesium concentration is usually reduced to 40 ppm.

Since the effluent from a softening unit is usually supersaturated with calcium carbonate at the usually high pH values, it is necessary to reduce the pH to a value that allows the solution to be exactly saturated for the calcium-ion and carbonate-ion concentrations present. The relationship is



The equilibrium expression is

$$K = \frac{[\text{Ca}^{2+}][\text{HCO}_3^-]}{[\text{H}^+]} = K_{\text{sp}} K_2 \tag{9}$$

or

$$\text{pH}_s = \text{p}K_2 - \text{p}K_{\text{sp}} + \text{pCa} + \text{p(alk)} \tag{10}$$

where pH_s is the pH of saturation, pK₂ is the second ionization constant for carbonic acid, pK_{sp} is the solubility product constant of CaCO₃, pCa is the negative logarithm of the molar concentration of calcium ion, and p(alk) is the negative logarithm of the bicarbonate ion concentration, which is measured analytically as the total titratable bases present at pH 6–9. The reduced pH is usually attained by recarbonation, ie, by the addition of CO₂. The carbon dioxide is obtained from the combustion of fuel oil or from liquid CO₂.

Modifications of the basic process are undersoftening, split recarbonation, and split treatment. In undersoftening, the pH is raised to 8.5–8.7 to remove only calcium. No recarbonation is required. Split recarbonation involves the use of two units in series. In the first or primary unit, the required lime and soda ash are added and the water is allowed to settle and is recarbonated just to pH 10.3, which is the minimum pH at which the carbonic species are present principally as the carbonate ion. The primary effluent then enters the second or secondary unit, where it contacts recycled sludge from the secondary unit resulting in the precipitation of almost pure calcium carbonate. The effluent settles, is recarbonated to the pH of saturation, and is filtered. The advantages over conventional treatment are reductions in lime, soda ash, and CO₂ requirements; very low alkalinities; and reduced maintenance costs because of the stability of the effluent. The main disadvantages are the necessity for very careful pH control and the requirement for twice the normal plant capacity.

The principal use for split treatment is for water with high magnesium content. This process also requires two units in series, which doubles the size of the plant. The ratio of the fraction of raw water in the primary unit to the fraction treated in the secondary unit is such that, when all of the magnesium is removed in the primary unit and none in the secondary, the mixed effluents contain 10 ppm magnesium as magnesium or 41 ppm as CaCO₃. For example, 75% of raw water containing 40 ppm magnesium would be treated in the primary unit and 25% in the secondary unit. Sufficient lime is added ahead of the primary unit to remove all of the magnesium or at least to an amount less than 1 ppm. This requires a pH of 11.1–11.3 with an excess of ca 70 ppm of caustic alkalinity. The effluent from the primary unit is mixed with the bypassed raw water in the secondary unit with no further addition of chemicals. The excess lime in the primary effluent is used to remove the calcium hardness from the bypassed raw water. In this process, the excess lime required for magnesium removal is used to soften raw water, whereas in conventional treatment this excess is wasted through recarbonation. Another advantage is the very low alkalinity (18–20 mg/L (68–76 mg/gal)) of the finished water. This, of course, reduces the necessary amount of soda ash.

One of the main problems associated with lime or lime-soda softening is the disposal of the sludge. Depending upon the ratio of calcium to magnesium removed and upon the amount of soda ash used, the sludge produced is 2.8–3.6 times the weight of the lime added. The principal methods of

disposal have been lagooning, discharge into the nearest water course, or discharge to the sanitary sewer system. The latter two are no longer acceptable methods of disposal because of the resultant pollution load.

A method of disposal for large plants is recalcination, ie, regeneration of lime from the CaCO₃ by heating the CaCO₃ in a kiln to remove the carbon dioxide. The lime can be reused in the plant, the CO₂ used for recarbonation, and the excess lime sold. In cases where Mg(OH)₂ has precipitated with CaCO₃, it is necessary to remove the magnesium prior to recalcination. This is accomplished by selectively dissolving the magnesium with CO₂ from the kiln at ca pH 10.3. At lower pH values, CaCO₃ dissolves and, at higher values, magnesium remains in the sludge. In some plants, the sludge is vacuum-filtered to ca 40–50% solids and then used for agricultural purposes or disposed of in a land fill.

Ion Exchange. For waters with high noncarbonate hardness or high magnesium content, or both, the use of lime or lime-soda may be very expensive. In such cases, ion-exchange softening may be more economical, particularly is brines are locally available at little cost for regeneration of the resins. Ion-exchange softening involves exchanging sodium or hydrogen ions for the calcium or magnesium ions in the raw water. The exchange reactions are equilibria that favor the formation of polyvalent ion-resin complexes over monovalent ion-resin complexes. The raw water to be softened is passed through the resins that are initially saturated with sodium or hydrogen ions. The exchange is referred to as the exhaustion reaction and may be shown as $Ca^{2+} + 2 NaR \rightarrow R_2 + 2 Na^+$, where R is the resin matrix. The regeneration step is simply the reverse of the exhaustion step with the concentration of the regenerant ion increased to 5–25% in order to reverse the reactions. For sodium-ion regeneration, sodium chloride is used and for hydrogen-ion regeneration, either sulfuric or hydrochloric acid is used. The theoretical requirement for 98 wt % NaCl to regenerate the resin is 453 g for each 686 g of hardness removed (expressed as CaCO₃). The actual requirements vary from 544 to 1088 g, depending upon the specific resin used and the regeneration conditions. The water passing through the exchanger is reduced to 0–2 mg/L (0–7.6 mg/gal) hardness. Since this is usually much lower than the final acceptable hardness, only a portion of the water is passed through the exchanger. Enough is bypassed to give the desired final hardness.

The resins used are highly cross-linked organic polymers with acidic functional groups. The most common of the resins used are sulfonated copolymers of styrene and divinylbenzene (see ION EXCHANGE).

There are several modifications of the ion-exchange process. For example, water with high total solids can be treated partially by sodium-cycle exchange followed by aeration to remove CO₂ and, consequently, part of the anions. However, sodium-cycle exchange increases the soluble-solids content as two sodiums (at wt % 23.0 – 46.0) are substituted for one calcium (40.0) or magnesium (24.3). Hydrogen-cycle exchange replaces one calcium or one magnesium with two hydrogens and, thus, reduces the total solids content. The disadvantage is the reduction in pH. The resultant water is very corrosive and the pH must be raised before distribution.

Another possible modification is the use of seawater as the regenerant. Even though it contains calcium and magnesium, but only 2.7 wt % NaCl, it sometimes can be purified by coagulation, filtration, and chlorination less expensively than salt can be purchased. The lower concentration reduces the regeneration efficiency by 40–50%.

The two types of installations used are fixed-bed and continuous-regeneration units. The continuous units consist of a closed circuit containing two sections, one for softening and one for regeneration; the resin is circulated countercurrent to the rawwater flow through the softening tank. The resin then is pulsed periodically into the regeneration tank, where it is regenerated and rinsed; then it is pulsed back to the softening tank. Thus, there is continuous softening without the downtime customary with fixed-bed units.

Taste and Odor Control. Tastes and odors in surface waters result from the action of biological organisms, eg, algae, or from various minerals, pollution by industry, domestic seepage, or agriculture. Groundwaters may have taste and odor if they are polluted or if they contain gases, eg, H₂S or CH₄; the latter always contains associated impurities that have taste and odor. Removal of these gases can be accomplished by adsorption (qv) with activated carbon (qv); oxidation with chlorine, potassium permanganate, or ozone; or aeration.

Organic materials are generally removed by addition of powdered activated carbon. The carbon may be added at any point in the plant, although it is advantageous to have as much contact as possible. The adsorption reaction is slow at room temperature, since it is diffusion-controlled. Oxidation with chlorine, potassium permanganate, or ozone may destroy tastes and odors or it may intensify them, depending upon the particular compounds involved. For example, chlorination of phenolic compounds leads to greatly increased tastes and odors. For this reason, the system must be studied in the laboratory prior to water treatment.

Hydrogen sulfide and methane can be removed by aeration, although the largest reduction in hydrogen sulfide may result from oxidation by the dissolved oxygen introduced during the aeration. At low pH values, the product is sulfate, whereas at high pH values, the product is free sulfur.

Control of Organic Compounds. There are two groups of organic compounds of concern: the trihalomethanes, which result from the use of chlorine as a disinfectant, and volatile organic solvents, which percolate through the soil thereby contaminating groundwaters. Trihalomethanes (THMs) result from the reaction of chlorine with natural organic precursors according to the general equation



The THMs of concern include chloroform, dichlorobromomethane, chlorodibromomethane, and bromoform. Control is based on the use of an alternative disinfectant, removal of precursors prior to chlorination, or removal of the THMs after formation. Use of alternative disinfectants has had varying degrees of success. The removal of precursors has been accomplished by optimizing coagulation, using activated carbon, or oxidizing with ozone, KMnO₄, or ClO₂. The removal of THMs after formation can be accomplished by air stripping or adsorption. In some cases, a combination of these processes must be used to reduce the THM concentration to less than the maximum contaminant level (MCL) of 0.10 mg/L (0.4 mg/gal). Some plants have had average concentrations as high as 2 mg/L (7.6 mg/gal).

Volatile organic contaminants occur primarily in groundwaters as a result of the disposal of industrial solvents on the ground or in soakage pits. The removal of these compounds has best been accomplished by the use of air stripping or adsorption on activated carbon.

Iron and Manganese Removal. Groundwaters or water withdrawn from the depths of reservoirs may contain soluble iron and manganese in the +2 oxidation state. Either one in equilibrium with dissolved oxygen exists in an oxidized and, therefore, insoluble state (Fe(OH)₃ and MnO₂). If the reduced metal ions are allowed to remain in the finished water and then come into contact with the atmosphere, the oxidized forms precipitate upon domestic fixtures or clothes, yielding a reddish-brown stain from the iron and a dark-brown-to-black stain from the manganese.

Both ions can be removed by oxidation and subsequent filtration. Aeration is adequate for iron(II) oxidation at above pH 6, but the oxidation of manganese(II) is much too slow, even at higher pH values, for effective removal. Potassium permanganate or chlorine dioxide is frequently used for the oxidation of manganese; however, their use must be followed by coagulation prior to filtration because of the formation of colloidal MnO₂.

BIBLIOGRAPHY

"Water (Municipal)" in the *Encyclopedia of Chemical Technology*, 1st ed., Vol. 14, pp. 946–962, by H. O. Halvorson, University of Illinois; "Water, Municipal Water Treatment" in *ECT* 2nd ed., Vol. 22, pp. 82–104, by J. E. Singley, University of Florida; in *ECT* 3rd ed., Vol. 24, pp. 385–406, by J. E. Singley, Environmental Science & Engineering, Inc.

1. S. T. Powell, *Water Conditioning for Industry*, McGraw-Hill Book Co., New York, 1954.
2. M. N. Baker, *The Quest for Pure Water*, American Water Works Association, New York, 1949.
3. J. Arboleda-Valencia, *Teoria, Diseno y Control de los Procesos de Clarificación del Agua*, Pan American Center for Sanitary Engineering and Environmental Sciences, Pan American Health Organization, Lima, Peru, 1973.
4. G. S. Logsdon and E. C. Lippy, *J. Am. Water Works Assoc.* **74**, 649 (1982).
5. **U.S. Pat. 293,740** (Feb. 19, 1884), I. S. Hyatt.

6.

T. R. Camp, *J. Am. Water Works Assoc.* **53**, 1478 (1961).

7.

W. R. Conley, *J. Am. Water Works Assoc.* **53**, 1473 (1961).

8.

G. Biederman quoted by K. Schlyter, *Trans. R. Inst. Technol. Stockholm*, 196 (1962).

9.

H. Leugweiler, W. Buser, and W. Feitknecht, *Helv. Chim. Acta* **44**, 805 (1961).

10.

J. L. Cleasby, *Proceedings, American Water Works Association Seminar*, No. 20153, American Water Works Association, Denver, Colo., 1980.

11.

V. K. LeMar and T. W. Healy, *J. Phys. Chem.* **67**, 2417 (1963).

12.

H. Kubota, thesis, University of Wisconsin, Madison, Wisc., 1956.

13.

Z. G. Szabo, L. J. Czanyi, and M. Kaval, *Z. Anal. Chem.* **147**, 401 (1955).

14.

J. E. Singley and A. P. Black, *J. Am. Water Works Assoc.* **59**, 1549 (1967).

15.

J. H. Sullivan and J. E. Singley, *J. Am. Water Works Assoc.* **60**, 1280 (1968).

16.

R. W. Ockershavsen, *J. Am. Water Works Assoc.* **57**, 309 (1965).

17.

A. P. Black, J. E. Singley, G. P. Whittle, and J. S. Maulding, *J. Am. Water Works Assoc.* **55**, 1347 (1963).

18.

W. Stumm and G. F. Lee, *Ind. Eng. Chem.* **53**, 143 (1961).

19.

J. Am. Water Works Assoc. **60**, 1317 (1968).

J. E. Singley

Environmental Science & Engineering, Inc.

SEWAGE

Sewage is the spent water supply of a community. Because of infiltration of groundwater into loose-jointed sewer pipes, the total amount of water treated may exceed the amount consumed. Sewage contains about 99.95% water and ca 0.05% waste material (1). The higher the per capita consumption of water, the weaker (more dilute) is the sewage, which may also be affected by industrial wastes. Sewage flow is greater during daylight hours and varies little in large cities, whereas in many small communities, the late-night flow is almost all groundwater.

Per capita flow varies from <378 L d (2) for a residential community to >1134 L d for highly industrialized areas. The concept of population equivalent is used for evaluating industrial waste contributions to sewage. It is applied when planning for hydraulic and BOD loadings.

Because of large equipment and land requirements, capital costs for wastewater-treatment plants are high. A collection system that conveys both sanitary and storm flows must be designed to deal with high peak flows at the treatment plant; detention basins are usually provided in order to smooth the flow into the plant and reduce the sudden peak flow. In the absence of such precautions, it may be necessary to by-pass a portion of the flow.

Some materials, such as gasoline and heavy metals, may damage the collection system, cause explosions, or upset the treatment processes. Therefore, they cannot be discharged directly into the system. These prohibitions are governed by sewer ordinances.

Strength of sewage is expressed in terms of BOD, total solids, suspended solids, fixed solids, volatile solids and filterable solids. The BOD is a measure of the load placed on the oxygen resources of the receiving waters. Treatment efficiency is usually evaluated on the basis of BOD removal by the plant. Unless otherwise stated, BOD signifies the biochemical oxygen demand for five days at 20°C (BOD₅). The significant parameters of wastewater are evaluated according to the analytical methods set forth by the U.S. Public Health Association (3).

Sewage Treatment and Disposal Systems

Sewage works were originally constructed for reasons based primarily on public health concepts. However, prevention of disease is not the sole purpose of modern water sanitation practice; it is also a means to protect the oxygen resources of the receiving water (2). An effluent that does not significantly reduce the oxygen concentration of the receiving water into which it is discharged can be expected to have a low concentration of food for microorganisms that deplete the dissolved oxygen. Thus, protection of the oxygen resources of the receiving water prevents the spreading of disease and satisfies aesthetic considerations.

Private and Rural Disposal Systems. In areas not served by sewers, human and other water-carried wastes are disposed of in primitive privies, cesspools, or septic tanks (4). In the more developed areas privies have almost disappeared. Cesspools are simply pits in the ground into which wastewater is allowed to flow, and in many parts of North America they are not permitted. Cesspool water seeps into the ground, leaving the solid matter in the pit, thus groundwater in the area might become contaminated. Septic tanks are widely used in smaller communities and outlying suburbs of larger communities. The tank is kept full of waste and functions as both a sedimentation tank and anaerobic digester. Sanitary and kitchen wastes flow into the tank; grease and light material rise to the top. Heavier materials settle to the bottom and decompose anaerobically. Baffles are placed at the inlet and outlet and a grease trap is usually provided in front of the tank, which has to be cleaned periodically of solids. Good practice requires a minimum volume of 5670 L (1500 gal). The effluent flows to a tile field where it is disposed of in the soil. The tile field is composed of perforated field tile fed by a manifold and underlain with granular material, usually gravel. Clogging of the soil under the tile field must be prevented. As more areas are served by municipal systems, septic tanks are becoming less common. Malfunctioning septic tanks cause odor problems and present a public health hazard. Furthermore, septic tanks and cesspools may form a closed system in which the waste discharged may return in the water supply. In many developing nations, human waste, called night soil, is not discarded, instead, it is used as crop fertilizer and for biogas production (5). Although it is not good public health practice, it is utilization of a valuable resource (see FUELS FROM BIOMASS; FUELS FROM WASTE).

The Imhoff tank is similar in many respects to the septic tank, operating as a sedimentation tank and digester. It is composed of two chambers, one above the other and the shape may be square, circular, or rectangular. The depth ranges from 7.7 to 10.6 m (6). Sewage flows slowly into the upper chamber at ca 0.3 m s and solids settle out and slide through a slot into the lower chamber, where the flowing wastewater is detained for ca 2 h. Solids held in the lower, or digestion, chamber have an initial water content of ca 95%. After a digestion time of 30–60 d the water content of the digesting sludge is reduced to ca 80% which greatly reduces the volume. The sludge is withdrawn at intervals for further dewatering and disposal. Gases produced during digestion are allowed to escape to the atmosphere through vents located at the tank sides. Solids buoyed up by gas are prevented from escaping to the upper chamber by deflector plates.

Small Communities. Small communities and recent subdivision additions to larger communities, which have not yet been connected to municipal collection systems, must have a means of waste disposal. Septic tanks are a possibility, but require periodic servicing and cleaning. Furthermore, the soil is not always suitable for accepting the effluent. An alternative is the package plant. These units are commercially produced to serve small areas. They furnish primary treatment and some secondary treatment, and require only minimal operating supervision. Capacity can be varied as needs dictate. In general, public health authorities prefer such installations instead of septic tanks.

City Disposal Systems. The systems described above are totally insufficient for population centers larger than small and rural communities. The operations necessary for treating sewage are outlined in Figure 1. In primary treatment large particles are removed by screening and sedimentation; 15 to 60% of the BOD can be removed in this manner. Colloidal and dissolved substances are removed by secondary, or biological treatment where the microorganisms of the process utilize the waste material for food. A commonly accepted definition of tertiary treatment is the use of any process or operation for further removals (2).

Operation and Maintenance. Engineering and construction firms can be contracted to take charge of treatment-plant operation and maintenance and manage other aspects necessary to meet performance goals, such as hiring and cost controls (7). Costs to the municipalities are reported to be competitive with the more traditional approach of management by municipal employees (8).

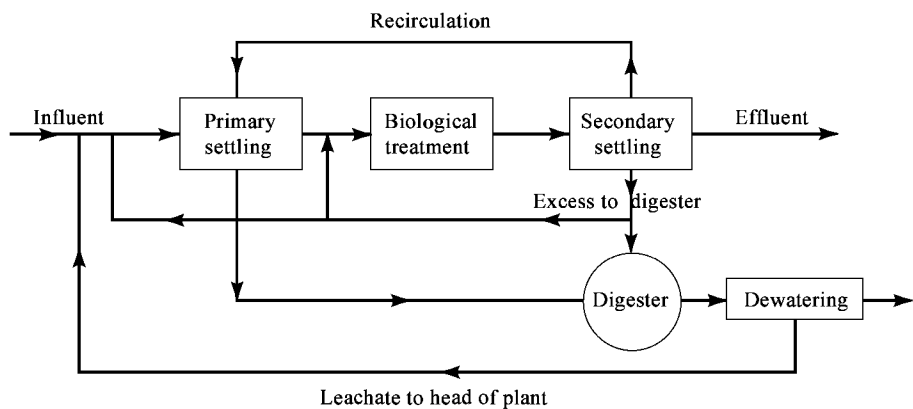


Fig. 1. Sewage-treatment operations.

Primary Treatment

Collection. It is desirable, whenever possible, to collect sewage by gravity flow (9). Otherwise pumping is necessary and flow velocities must be maintained above 0.6 m s in order to prevent solids from settling out in the pipes. Low flow velocities, long detention times, and relatively high temperatures cause operating problems in many areas. Sewage is best treated if it arrives at the treatment plant in a fresh condition. The plant usually has multiple treatment units and treatment continues without interruption during periods of routine maintenance or repairs.

The pumps must be protected against large objects (wood, etc) in the flowing sewage. Coarse racks with clear openings >5 cm are placed in advance of grit chambers. Settling tanks should have clear openings of 2.5–5 cm which must be cleaned frequently (generally by hand) to prevent excessive hydraulic head loss. Raking are usually buried, incinerated, digested with sludge, or composted.

Mechanically cleaned racks allow smaller clear openings because head loss does not become so high. Mechanical cleaning can be intermittent or continuous. Intermittent cleaning is cycled by float-operated switches controlled by a float in the influent channel.

Comminutors macerate floating material into particles too small (0.5–1 cm) to clog pumps. Comminutors have almost completely replaced racks and screens with small openings.

Grit Chambers. These are placed ahead of sedimentation chambers and are designed to remove inorganic suspended matter of ca 0.02 cm diameter, but not organic matter. Removal of grit provides protection against excessive wear of pumps and loss of volumetric capacity of sludge digesters. Large digesters in plants serving low-lying sandy areas can lose as much as one third of their capacity in a few years. Organic matter in the flowing wastewater is deposited when the flow velocity is below 23 cm s. The mixture of grit and organic material is called detritus and is unsuitable for landfilling. Velocities above 38 cm s cause the smaller particles of grit to be resuspended (10). Thus, a fairly narrow velocity range is required. The velocity is controlled by means of a Parshall flume placed immediately after the grit chamber. These flumes, sometimes called open-channel venturis, operate on the principle of critical flow (11) and measure the volumetric flow rate. However, they also control the velocity of flow in the grit chamber. Grit chambers often have a parabolic shape because flow velocity in the flume follows a parabolic function with water depth. The amount of grit removed per $3.78 \times 10^3 \text{ m}^3$ (10 gal) is 29–339 L (7.7–89.6 gal). In smaller plants, grit is removed periodically by hand, whereas in larger plants an automatic conveyor system is used.

Sedimentation. In the sedimentation (settling) tanks, solids settle and oil and grease rise to the surface where they can be skimmed and transferred to the sludge digester. Tanks may be circular or rectangular and depths are 2.1–4.5 m. Current practice favors the circular configuration. The bottom is sloped at ca 1° in rectangular tanks and ca 8° in circular tanks to facilitate sludge removal. Detention periods are 2–6 h. Design is on the basis of application rate, usually 0.472 L (m²s) [1000 gal (yd²·d)] for primary settling tanks and 0.283–0.377 L (m²s) [600–800 gal (yd²·d)] for secondary (humus) tanks.

Secondary Treatment

The two main processes utilized for secondary treatment are the trickling filter and activated sludge (12). Existing biological-treatment methods are logical developments from sewage farms (irrigation areas) to intermittent sand filters to contact (fill and draw) beds. Numerous modifications of the basic processes have evolved over the years but the underlying principles remain unchanged (13).

Modifications have been developed in response to specific operating needs. The basis of biological treatment is the formation of an environment in which the microorganisms can thrive under controlled conditions (14). The microorganisms originate in the sewage. A suitable environment is rich in food and maintained in an aerobic state. It has been said that a sewage-treatment plant is a river in miniature (14).

Irrigation by sewage (sewage farms) provides water for irrigation and some sewage stabilization. This disposal method is still applied in some arid regions but is not desirable from a public health viewpoint. It cannot be used when the applied sewage comes in contact with crops. It may be used for orchards but with a certain risk. A notable case involved sewage contamination of ripening olives; odors cannot be avoided.

Intermittent sand filters resemble slow sand filters used for the treatment of potable water. The sewage is applied to a sandy area and allowed to flow slowly downward. Raw sewage can be applied at rates as high as 8.4 L (has) [0.9 gal (acres)] and as high as 84.4 L(has) [9 gal (acres)] after biological treatment. The latter would be considered a tertiary treatment step. Surface accumulations are periodically removed. This method may not be applied to areas on top of extensive limestone formations because of the danger to active wells. Biological films that form on the sand grains undergo continuous stabilization. Beds must be rested between dosings.

Fill-and-draw beds operate as the name indicates. A tank is filled with sewage and allowed to remain full for a while. It is then drained and allowed to rest. Air is drawn into the bed during emptying. Bed loadings are ca 21.1 L (has) [2.25 gal (acres)]. This method is used more for industrial than for municipal wastes.

Trickling Filters. The so-called trickling filter is not a filter but a bed of stones or other coarse material (packing) over which the sewage flows. In terms of the total number of installations, it is the most widely used biological treatment process. However, the greatest total volume of waste is treated by the activity-sludge process (12).

Settled sewage is distributed by slowly rotating arms equipped with nozzles and deflectors and is allowed to flow downward over the packing. Air is drawn into the filter by the temperature differential between the inside and outside of the filter, thus maintaining a supply of oxygen. Until recently, stones with diameters of 2.5–10 cm were used as filter medium. These stones permitted sufficiently loose packing to allow free flow of water and air and left sufficient openings to prevent clogging by biological slimes. Stone filter depth is 0.9–4.2 m, and is usually 1.8 m. Plastic filter media are being used more and more (15). They require filters that are much deeper than those using stones and pumping costs can be expected to be high over the life of the plant. However, higher removal efficiencies are reported.

Sewage flows slowly downward over the filter medium and the effluent is collected in vitrified tile underdrains which collect the filter effluent and circulate air into the filter. The underdrains discharge into a main collection channel which, in turn, discharges into the final settling tank.

The importance of the final settling, or humus, tank can be seen by an examination of what occurs in the trickling filter itself. A new filter is broken in by applying settled sewage as in the normal operation. After a period of time the microbial, or zoogeal, mass forms on the filter medium and stabilizes the waste. Waste material is first adsorbed, and then assimilated by the microorganisms.

Much of the waste material is transferred to the microbial cells and is utilized for cell synthesis and energy (16). Dead and dying organisms are separated from the zooglear mass and collected in settling tanks to prevent the clogging of the filter. A rate of application lower than that of the primary tanks is necessary because the organisms to be removed are considerably lighter than the material removed in primary tanks; the application rate is usually 28 L (m²s) [18.6 gal (yd²s)]. Trickling filters may be classified on the bases of hydraulic loading per unit area per unit time and the amount of BOD per cubic meter per day.

Low Rate. The low rate filter is most commonly used in small plants. The loadings are 0.22–0.43 m³ (has) [23.5–46 gal (acres)] and 0.16–0.32 kg BOD (m³s) [0.6–1.2 g BOD (gass)] (7,17). With proper operation, 80–85° of BOD is removed. Generally, appreciable amounts of nitrite and nitrate are found in the effluents, indicating a high degree of organic stabilization.

A portion of the effluent is recirculated, in order to smooth out flow, keep the food concentration constant, lower film thickness and control psychoda flies, and reseed the applied sewage with acclimatized organisms. The psychoda, or filter fly is a very small insect that breeds in thick trickling-filter slimes. It does not bite, but can be a nuisance. Its radius of flight is small, but it can be carried great distances by the wind. The fly can be controlled in the development phase by occasional flooding of the filter or chlorination of the applied sewage.

Intermediate Rate. At 0.43–1.08 m³ (has) [46–115 gal (acres)], certain operational difficulties were frequently encountered and therefore this range was avoided for many years. It now appears that these difficulties stemmed primarily from the inability of the hydraulic load to keep the filter slimes from becoming too thick and thereby clogging the filter. The use of relatively large filter stones solves this problem.

High Rate. By increasing the hydraulic loading of sewage to 1.08 m³ (has) [115 gal (acres)], BOD loadings per unit volume of filter were increased, which, in turn, increased the organic material in the effluent. High recirculation ratios maintained the high hydraulic loading, and at the same time reduced the organic concentration in the effluent. Hydraulic loadings are 1.08–4.32 m³ (has) [115–462 gal (acres)] and up to 1.44 kg BOD (m³s) [5.5 g BOD (gals)], but removal efficiency is only 65–75°.

Superrate. High BOD removal efficiencies (97°) at hydraulic loadings of 10.8 m³ (has) [1150 gal (acred)] were obtained in experimental plants with plastic media. In these plants, much of the microbial mass remains in the recirculated effluent. This process is, in effect, a modification of activated sludge. Organic loadings are ca 1.60 kg BOD (m³s) [6.1 g BOD (gals)].

Modifications. Modifications of the trickling-filter included improvement in media, air circulation, loading, and means of bringing the waste into contact with the zooglear mass. In one modification, sewage was introduced at various levels of a very deep filter in order to distribute the load more evenly. This can be compared to the step–aeration modification of the activated sludge process. Hydraulic loadings of this modification are as high as 54 m³ (has) [5800 gal (acres)].

Rotating Biological Contactors. Rotating biological contactors (RBC) (18) are composed of multiple plastic or stainless steel disks which rotate slowly through incoming settled wastewater; ca 40° of the disk area is submerged. A growth similar to trickling-filter slime develops on the contactor surface and oxygen is supplied from the air when the disk is exposed. Recirculation of effluent is not necessary. The principal design parameter is the wastewater flow rate per unit disk area, ca 155 L (m²d) [103 gal (yd²d)]. Peripheral speeds are ca 18 m min. Using four stages it is possible to obtain an overall BOD reduction of 90°. The RBCs, compared to activated sludge, have low energy requirements, resist shock loads, and produce a highly nitrified effluent.

Activated Sludge. In this process, in contrast to trickling filters, activated sludge floc is suspended in the moving wastewater stream. It originated in efforts to purify sewage by blowing air into it. It was observed that flocs composed of voraciously feeding organisms developed after prolonged aeration. After aeration was stopped, the floc settled. Addition of fresh sewage to the tank containing the sludge produced high purification in a reasonable time, and therefore, the floc was called activated sludge. As in the case of the trickling filter, it was first operated as a fill-and-draw bed. Further research proved that continuous operation could be practiced.

Some activated sludge is returned to the aeration tank influent and the excess sludge discharged to digestion. The sludge–sewage mixture is then purified by aeration, and the aeration tank effluent is settled to remove floc from the plant effluent. The last step removes the organic material transferred from the flowing wastewater to the cell mass. The floc is formed in sewage by aerobic growth of unicellular and filamentous bacteria, eg, protozoa and other organisms. These organisms gain food and energy by feeding upon the sewage.

Aeration tanks are usually rectangular in cross section, 3–4.5 m deep and 9 m wide. The ratio of surface length to width should be >5 : 1 to ensure thorough mixing. Detention periods in the aeration tanks are 4–8 h. This is a strictly aerobic process and air requirements are high. Satisfaction of oxygen requirements is difficult because oxygen is sparingly soluble in water (10 mg L or 38 mg gal).

For aeration, diffused-air or mechanical units are used. Air diffusers are commonly used in North America, but some mechanical aeration units are installed in plants with capacities >3,800 m³ d (10 gal d). Aeration transfers oxygen to the sewage and maintains aerobic conditions, mixes sewage and floc, and keeps the floc in suspension.

The air introduced from the diffusers creates a spiral flow pattern, thus promoting the mixing of floc and sewage and eliminating dead spaces. Rotating brushes, surface impellers, or submerged vertical turbine aerators provide mechanical aeration. Rotating brushes and mechanical surface aerators help the transfer of oxygen from the air at the air–water interface. Draft tubes are required for deep tanks. Slow speed turbines, located ca 0.5 m above the bottom of the aeration tank, distribute compressed air throughout the tank. They can be used with deep, narrow tanks, but compressed air is required.

The floc returned to the aeration tank (10–50°) has the same function as trickling-filter slime. However, the floc concentration can be varied as operational needs dictate. Usually the so-called mixed liquor–suspended solids (MLSS) are 600–4000 mg L (2.3–15 g gal).

A very important parameter in routine control of the process is the ratio of the MLSS volume to the dry weight, called the Sludge Volume Index (SVI). A well operating plant has an SVI of 50 to 100, but operational difficulties occur when the SVI approaches 150. At this point the sludge settles poorly.

The activated-sludge process is microbiological in nature and factors that promote or inhibit growth are important, including the pH, temperature, and oxidation–reduction potential. The pH determines which organisms predominate in the system. Bacteria are the dominant organisms at pH 6.5–9, fungi below pH 6.5. In the presence of metabolic products, buffering capacity must be adequate.

Modifications are usually the result of particular operating needs. Future improvements can be expected by process development rather than changes in hardware (19).

High Rate Activated Sludge. A low MLSS concentration of 200–500 mg L (0.8–1.9 g gal) is maintained (15), resulting in a high ratio of food to microbial mass, which keeps the floc in an active growth phase; excess food is discharged in the effluent. Although the BOD is not removed at a high rate, there are areas in which this technique is adequate. New York City uses this method because the sewage is weak and temperatures are low. Poor results were obtained in Philadelphia and Los Angeles because the sewage is strong, temperatures are higher, and pipe detention times are long (16).

Completely Mixed Activated Sludge. Intimate mixing of returned sludge and settled wastewater, homogenizes the contents of the aeration tank, including the organic load. The wastewater is thus subjected immediately to attack by fully developed organisms; biological stability is enhanced by this design (20).

The net growth rate of the microbial population approaches zero by increasing the aeration time in order to extend the period in which the activated sludge is in the endogenous-growth phase. Although it is not possible to operate an extended aeration system without sludge accumulation, the resulting sludge is inert to activated-sludge organisms.

Tapered and Step Aeration. In New York City, plants scattered throughout the five boroughs are treating more than 3.8 × 10⁶ m³ (10⁹ gal) of sewage each day. In conventional plants, sewage is introduced at one end of the aeration tank and flows straight through. The oxygen demand at the influent end exceeds that at points farther along the flow path. In order to satisfy the oxygen requirements, the number of diffusers was increased at the beginning and decreased toward the effluent end. This arrangement is known as tapered aeration and is now common in conventional plants. When all of the floc of sewage is added at the beginning of the aeration tank, the oxygen demand is high because of the large supply of microbial food. In step aeration, the sewage (food) is fed at various intermediate points along the aeration tank. The microbiological food is thus applied evenly and oxygen requirements are lower. This modification has increased the capacity of existing overloaded plants.

Biosorption. This process makes use of a phenomenon observed but ignored by many investigators. If activated sludge floc and raw sewage are mixed and aerated, there is first a significant reduction in BOD of the effluent. It is followed by a notable BOD rise and then by a decrease. Originally,

the initial decrease had been attributed to experimental error; however, it was shown to be due to adsorption of colloidal material onto the activated sludge floc. Plants serving Austin, Texas, and Bergen County, New Jersey, were converted from being overloaded to underloaded by changing from conventional activated sludge to biosorption.

Kraus Process. The municipal wastewater-treatment plant at Peoria, Illinois, received a waste high in carbohydrates, causing a nitrogen and oxygen deficiency in the activated sludge which was light and settled poorly. Addition of a sludge-digester nitrogenous supernatant to the plant influent corrected the imbalance and the resulting floc settled well. Removal of organic material from the flowing wastewater was improved (16).

Sludge Digestion and Disposal

Putrescible material collected from primary and secondary tanks must be disposed of cheaply and efficiently because it is a nuisance source. It can be stabilized by biological means. Raw sludge is ca 95% water and the 5% solids cannot be easily separated. As sludge is broken down, the water content is lowered and the volume is greatly reduced. Water content of sludge reduced from 95 to 90% reduces the sludge volume by one-half and the reduction to 85% reduces the volume to one-third of the original. Fresh sludge has a gray color and can be easily pumped. It has a most disagreeable odor, due primarily to thiols.

Digestion. Digestion reduces the volume and the pathogenic organisms. Digested sludge is black and granular and has a tarry color. Sludge is withdrawn periodically from the settling tanks and allowed to flow by gravity to a collection well. From there it is pumped to a digester where it is thoroughly mixed.

Sludge digestion can be aerobic or anaerobic. The equipment used for aerobic digestion closely resembles that used for the activated-sludge process (18). Aerobic digestion is mostly applied to waste-activated sludge. Most of these digesters are for small-to-medium plants and most package plants have aerobic digesters. Aerobic digesters are operated as either batch or continuous-flow reactors; the latter type is more common.

Sludge is destroyed by microorganisms and the kinetics of their life processes is temperature dependent. Short anaerobic digestion detention times are obtained at 35°C. Even shorter detention times are possible at 52–54°C, but detention in this range is costly. An increase in detention time occurs at 35–43°C and then a progressive decrease takes place until 52–54°C. This variation is caused by a change in character of the dominant process organisms.

In older installations, the digesting sludge was heated by hot-water coils in the tank periphery. It was thought that adequate mixing resulted from turbulence induced by gases generated, however, this was not the case. Heat was transferred only in the immediate vicinity of the coils, and the remainder of the tank was not heated well. External heat, fueled by sludge digestion gases, results in better mixing and more uniform temperatures. External heat exchangers have almost completely replaced hot-water coils in new plant construction.

Sludge gas consists of 65–70% methane and 25–30% carbon dioxide. Anaerobic sludge digestion also produces traces of H₂S and hydrogen. Generally, the gases generated in sludge digestion are sufficient to provide for the immediate fuel needs of the plant in order to maintain the digester temperature, provide hot water, and fuel for incineration of sludge (if practiced) and generators. Electricity constitutes the principal operating cost of a treatment plant, but significant savings may be possible with on-site generation. In any case, stand-by generation must be provided in case of power failure.

Volatile acids, reported as acetic acid, are the most important operational parameter. In a well-operating digestion process, the value should be <1 g/L (3.8 g/gal). A value >6 g/L (23 g/gal) indicates malfunctioning; optimum pH is 6.8–7.2, and a pH <6.8 indicates excessive volatile acid production. Formerly, lime was added to the digester contents if the pH showed an undesirable drop. However, the reduction in pH indicated a change in organism that could not be remedied with lime (2).

Aerobic digestion causes fewer maintenance problems than anaerobic digestion; capital costs are lower, and supernatant BOD is lower. The advantages of anaerobic digestion are lower energy requirements because mixing and aeration are not necessary and the production of methane is a useful by-product. Furthermore, the solids content of the digested sludge is higher and, therefore, the volume of sludge to be handled significantly lower. Addition of a thickener-clarifier before aerobic digestion greatly reduces this difference.

Disposal. Digested sludge is reasonably inert but has a high water content. It can be dewatered by filtration or by drying on open or covered sand beds. Open sand beds are exposed to the atmosphere and the sludge dries by downward drainage and evaporation. A covered bed resembles a greenhouse with high temperatures and humidity. In both cases, sludge is allowed to flow by gravity from the digester to the sand beds and then permitted to stand for the required time. The dried sludge is scraped from the beds and taken to ultimate disposal. Sludge may also be dewatered by vacuum filtration.

The dried sludge may be incinerated or used for landfill or fertilizer. Sewage sludge does not have wide application as a fertilizer (qv), partly for aesthetic considerations and partly because of the low nutritive values. Refusal by the public to accept sewage sludge as fertilizer came long before recognition of the danger of toxin accumulation.

Many substances that are separated from the flowing wastewater may appear essentially unchanged in the final digested sludge. Concentrations of undesirable materials may be greatly increased in the sludge. The ultimate disposal site must be chosen with great care and contained in order to prevent release of toxic substances to the surrounding environment. Toxic materials may concentrate through biomagnification as they proceed up the food chain.

Ocean Disposal. Disposal of raw or treated sludge by barging to sea was practiced for many years by some coastal cities, but today is highly controversial, and it appears that this method will be no longer economically feasible. Federal regulations require that sludge be taken to disposal sites about 160 km from the coast, whereas formerly, disposal sites were permitted within 20 km offshore. Transportation costs are expected to be so high that ocean disposal will be discontinued.

Chlorination of Effluent

It is common practice to chlorinate wastewater effluents for bacterial control. Regulations vary from state to state, but all require chlorination to specified residual concentrations. Regulations change with the season and the most stringent are in effect during the swimming season. A phenomenon not fully understood is aftergrowth. Immediately after discharge to the receiving water, the bacterial count is low, but then suddenly rises. Questions have been raised concerning the benefit of effluent chlorination. It has been found that some unusual organic industrial chemicals pass unchanged through conventional treatment plants. Chlorination of these substances gives products that are suspected of being carcinogenic. Drinking such waters may be dangerous to health. These considerations have brought about renewed interest in ozone, other halogens, and ultraviolet irradiation as means of bacterial control.

Stormwater overflow is chlorinated, and in some plants the influent may be chlorinated for odor control.

Other Disposal Problems

Watercraft-Waste Disposal. The popularity of recreational boating has brought with it the problem of disposing of wastes generated by the boat users. In many rivers and lakes, this has become a serious problem. A large number of pleasure craft can place, in a weekend, a pollutional load equivalent to a small community on a medium-sized water body. As legislation was passed by the states, it became apparent that this problem would require a solution at the national level. Various technologies are available, but no agreement could be reached on which to use. Units available include the holding tank, macerator–chlorinator, and on-board incinerators. The Coast Guard was given the task of recommending a solution but became mired in bureaucratic tugs-of-war. It was agreed that holding tanks would keep the waste from reaching the receiving waters directly from the boat. However, availability of facilities to receive the retained waste material ashore was limited. In many cases, the pumped-out waste was discharged to the receiving water as soon as the sun had set. After several years of indecision, rules were promulgated that allowed the use of the macerator–chlorinator for a period, followed by total conversion to holding tanks. Macerator–chlorinators chop the waste into fine particles for adequate chlorine contact and are regarded as potential sources of bacterial contamination. Incinerators did not achieve the same acceptance on pleasure craft as they did on railroads.

Surfactants. Long-chain soaps degrade using biological treatment and digestion (see SOAP). In the early 1950s, synthetic detergents (syndets) came into wide use; they are mainly based on alkylbenzenesulfonates (ABS) with phosphates as builders (see SURFACTANTS AND DETERSIVE SYSTEMS). The early syndets were resistant to biodegradation and passed unchanged through wastewater-treatment processes. Excessive foam formed in activated-sludge

aeration tanks, points of high turbulence, and receiving waters. Suds production was not essential for cleaning efficiency, but was a merchandizing asset. Producers were reluctant to replace high suds detergents with a low suds product. In Europe, sewer ordinances were enacted, which forbid discharge of biologically hard detergents after a grace period. In North America the approach was that of gentle persuasion. In most countries, substitution of high foaming by biodegradable detergents was accepted as the best solution.

Phosphates. Phosphates are used as builders in detergents and function primarily as water softeners. A critical NO₃(N):PO³⁻₄(P) ratio of 1:15 is necessary before algal blooms occur, which are essential for eutrophication (see WATER POLLUTION). It is necessary to reduce the amount of phosphate entering a treatment plant, which phosphates pass unchanged. The total phosphate entering the sewer systems has been greatly reduced by introduction of low phosphate detergents. In the period of controversy over lowering phosphates it was also suggested that highly nitrified effluents might be undesirable.

Health and Safety Factors

Wastewater-treatment plants have numerous hazards to be expected in a chemical-process plant. Worker safety is covered by applicable OSHA, state and local standards, but two hazards require special notice (9); the potential for infection by pathogenic organisms is always present, and plant workers require inoculation against the common waterborne diseases. In addition, since wastewater treatment plants utilize deep water-filled tanks, provision must be made against drowning. Proper railings should be installed around the tanks and life-saving equipment should be available for immediate use. A special hazard is the biomass concentration in aeration tanks. Ingestion of this floc has caused a number of fatalities.

Government Regulations

The modern approach to wastewater treatment, protection of the oxygen resources of the receiving waters, requires that all aspects of the problem be addressed, ie, the systems approach. The Ohio River Sanitation Commission (ORSANCO) is an excellent example of basin-wide management dealing with situations that involve several political entities. This approach has been adopted in several other regions.

A succession of federal agencies and administrations has been charged with dealing with wastewater. At present this responsibility resides with the EPA. It has been proposed that most of the EPA functions be turned back to the states. It remains to be seen if the states, with conflicting needs and priorities, will be able to deal successfully with these problems.

The National Pollutant Discharge Elimination System (NPDES) is a cornerstone of the federal efforts to control water pollution (21). It determines what can be discharged to a publicly owned treatment plant. Indirect discharges may not be required to obtain an NPDES permit but must meet pretreatment effluent limitations and conditions of the NPDES permit of the treatment plant cannot be exceeded.

Applicable technologies to meet present guidelines and standards for wastewater treatment are given in Table 1.

Other Developments

Some wastewater-treatment plants have utilized pure oxygen in place of air. This permits higher oxygen concentration gradients within the liquid phase and thus allows higher biomass concentrations (24). The use of pure oxygen requires less energy, but the cost of pure oxygen must be taken into account. The aeration tank is under slight pressure and three stages are usually provided. Mixed liquor-suspended solids are 3–10 g/L (11–38 g/gal). Significant increases in volumetric loading rates, reduction in sludge production, and lowered treatment costs are reported.

Table 1. Applicable Technologies to Meet Present Guidelines and Standards

Process industry	Effluent guidelines and standards ^a		Applicable treatment technologies		
	Reference ^a	Parameters ^b regulated	Neutralization	Sedimentation and or filtration	Biological treatment
dairy products	405	pH, BOD, TSS		X	X
grain mills	406	pH, BOD, TSS		X	X
canned preserved fruits and vege-tables processing	407	pH, BOD, TSS		X	X
canned preserved seafood	408	pH, BOD, TSS O and G		X	X
pro-cessing sugar processing	409	pH, BOD, TSS, O and G, tempera-ture, fecal coliform	X	X	X
textile mills	410	pH, BOD, TSS, O and G, COD, sul-fide, phenol, chromium	X	X	
cement manu-facturing	411	pH, TSS, tempera-ture	X	X	
feedlots	412	BOD, fecal coliform			X
electroplating	413	pH, TSS, metals, cyanide, total toxic organics (TTO)	X	X	
organic chemicals, plastics, and synthetic fibers	414	pH, BOD, TSS, metals, cyanide, volatiles, semi-volatiles	X	X	X
inorganic chemicals manufacturing	415	pH, TSS, COD, O and G, metals, chlorine, TOC, ammonia, cyanide	X	X	
soaps and detergents	417	pH, BOD, TSS, O and G, COD, sur-factants (MBAS)	X	X	X
fertilizer manu-facturing	418	pH, BOD, TSS, phos-phorus, ammonia, organic nitrogen, nitrate, fluoride	X	X	
petroleum and petroleum re-fining	419	pH, BOD, TSS, O and G, COD, sul-fide, chromium, hexavalent chro-mium, phenolic compounds (4AAP), ammonia	X	X	X

Table 1. (Continued)

Applicable treatment technologies							
Chemical precipitation	Chemical oxidation re duction	Air and or steam stripping	Activated carbon adsorption	Resin adsorption	Ion exchange	Ultrafiltra-tion and or reverse osmosis	Floatation ph ase separation
	X						X
X	X						X
X	X				X	X	
X	X	X	X	X	X	X	
X	X	X			X	X	X
X	X					X	X
X	X	X			X		
X	X	X	X	X	X		X

Table 1. (Continued)

Process industry	Effluent guidelines and standards ^a		Applicable treatment technologies		
	Reference ^a	Parameters ^b regulated	Neutralization	Sedimentation and or filtration	Biological treatment
iron steel manu-facturing	420	pH, TSS, O and G, metals, cyanide, phenols (4AAP), ammonia, naphtha-lene, benzene, tet-rachloroethylene, benzo(a)pyrene, chlorine	X	X	
nonferrous metals	421	pH, TSS, O and G, benzo(a)pyrene, metals, cyanide, ammonia	X	X	
phosphate manu-facturing	422	pH, TSS, phospho-rus, fluoride	X	X	
steam electric power gener-ating	423	TSS, O and G, metals, chlorine, priority pollutants	X	X	
ferroalloy manu-facturing	424	pH, TSS, metals, cyanide, phenols, ammonia	X	X	
leather tanning and finishing	425	pH, BOD, TSS, O and G, chro-mium, sulfide	X	X	X
glass manufac-turing	426	pH, BOD, TSS, Oil, COD, lead, fluor-ide, phenol, phos-phorus, ammonia	X	X	
asbestos manu-facturing	427	pH, TSS, COD	X	X	
rubber processing	428	pH, BOD, TSS, O and G, COD, metals	X	X	X
timber products	429	pH, BOD, TSS, O and G, phenols, chromium, arsenic, copper	X	X	X
pulp, paper, and paper board mills	430	pH, TSS, BOD, zinc, pentachlo-rophenol	X	X	X
builder's paper and board mills	431	pH, BOD, TSS, settleable solids, pentachloro-phenol	X	X	X

Table 1. (Continued)

Applicable treatment technologies							
Chemical precipitation	Chemical oxidation re duction	Air and or steam stripping	Activated carbon adsorption	Resin adsorption	Ion exchange	Ultrafiltra-tion and or reverse osmosis	Floatation ph ase separation
X	X	X			X	X	X
X	X	X			X	X	X
X					X		X
X	X		X		X	X	X
X	X	X			X	X	
X	X						X
X	X	X			X	X	X
X							X
X	X		X				
X			X				

Table 1. (Continued)

Applicable treatment technologies							
-----------------------------------	--	--	--	--	--	--	--

Process industry	Effluent guidelines and standards ^a				
	Reference ^a	Parameters ^b regulated	Neutralization	Sedimentation and or filtration	Biological treatment
meat products	432	pH, BOD, TSS, O and G, fecal coliform, am-moniam	X	X	X
metal finishing	433	pH, TSS, TTO, metals, cyanide	X	X	
coal mining	434	pH, TSS, iron, manganese	X	X	
offshore oil and gas extraction	435	O and G			
mineral mining and processing	436	pH, TSS, fluoride	X	X	
pharmaceutical manufacturing	439	pH, BOD, TSS, COD, cyanide, TSS, settleable solids	X	X	X
ore mining and dressing	440	pH, TSS, settleable solids, metals, ammonia	X	X	
paving and roofing materials	443	pH, BOD, TSS, O and G	X	X	
paint formulating	446	no process water discharge			
ink formulating	447	no process water discharge			
gum and wood chemicals manufacturing	454	pH, BOD, TSS	X	X	X
pesticide chemicals manufacturing	455, pending	pH, BOD, TSS, COD, organic pesticide chemicals	X	X	
explosives manufacturing	457	pH, BOD, TSS, COD, O and G	X	X	
carbon black manufacturing	458	O and G		X	
photographic processing	459	pH, cyanide, silver	X	X	
hospitals	460	pH, BOD, TSS	X	X	X
battery manufacturing	461	pH, TSS, COD, O and G, metals	X	X	
plastics molding and forming	463	pH, BOD, TSS, O and G	X	X	X
metal molding and casting	464	pH, TSS, O and G, TTO, phenols, copper, zinc, lead	X	X	

Table 1. (Continued)

Applicable treatment technologies							
Chemical precipitation	Chemical oxidation reduction	Air and or steam stripping	Activated carbon adsorption	Resin adsorption	Ion exchange	Ultrafiltration and or reverse osmosis	Floatation phase separation
							X
X	X				X		
X					X	X	
X	X		X	X	X	X	X
							X
						X	
			X	X	X	X	
			X	X		X	X
X	X				X	X	X
X					X	X	X
X	X				X	X	X

Table 1. (Continued)

Process industry	Effluent guidelines and standards ^a		Applicable treatment technologies		
	Reference ^a	Parameters ^b regulated	Neutralization	Sedimentation and or filtration	Biological treatment
coil coating	465	pH, TSS, O and G, metals, cyanide, TTO	X	X	
porcelain enameling	466	pH, TSS, O and G, metals	X	X	
aluminum forming	467	pH, TSS, O and G, metals, cyanide, TTO	X	X	
copper forming	468	pH, TSS, O and G, metals, TTO	X	X	
electrical and electronic components	469	pH, TSS, metals, TTO	X	X	

nonferrous metals forming and metal powders	471	pH, TSS, O and G, metals, cyanide	X	X	
waste treatment	437, pending	proposal pending	X	X	X
metal products and machinery	438, pending	proposal pending	X	X	
industrial laundries	441, pending	proposal pending	X	X	
transportation equipment cleaning	442, pending	proposal pending	X	X	

^a Refs. 22, 23.

^b BOD Biological oxygen demand. COD Chemical oxygen demand. TSS Total suspended solids. O and G oil and grease.

Table 1. (Continued)

Applicable treatment technologies							
Chemical precipitation	Chemical oxidation re duction	Air and or steam stripping	Activated carbon adsorption	Resin adsorption	Ion exchange	Ultrafiltra-tion and or reverse osmosis	Floatation ph ase separation
X	X				X	X	X
X					X	X	X
X	X				X	X	X
X	X				X	X	X
X	X				X	X	X
X	X	X	X	X	X	X	X
X	X				X	X	X
X					X	X	X

High pressure application is under investigation (25). The vertical tube reactor (VTR) uses an extended vertical U-tube with concentric tubes for achieving elevated pressures and temperatures and longer retention time for wet chemical reaction. The tubes may extend as much as 1.5 km into the earth. The waste is pumped into the inner tube along with air. As the waste stream and air flow downward the stream undergoes pressurization due to hydrostatic head. Heat is generated by the exothermic reactions of the chemical oxygen demand (COD). The process is reported to produce a stabilized effluent in the outer tube and spares land area.

The fluidized-bed bioreactor (FBBT) (26) increases the capacity of existing plants. Primary effluent is passed upward through the columnar reactor filled with sand or carbon with sufficient velocity to fluidize the bed. An attached biomass develops on the bed particles. Intimate contact between the biomass and waste is provided and improved removals are reported. Oxygen is provided by a deep U-tube reactor. No biomass recirculation is required and a secondary clarifier is not necessary.

Another process for sludge stabilization and volume reduction operates on the principle of superoxidation. In the oxyozosynthesis process, small amounts of ozone are applied to acidified sludge in a pressurized vessel. The reactor contents are maintained in a highly turbulent state. The product is reported to have the consistency of coarse paper and be biologically stable. Though promising, no explanations have been advanced for process dynamics and no cost figures have been given (27).

BIBLIOGRAPHY

"Sewage" in *ECT* 1st ed., Vol. 12, pp. 191–207, A. M. Buswell, Illinois State Water Survey Division; "Water, Sewage" in *ECT* 2nd ed., Vol. 22, pp. 104–115, by J. R. Pfafflin, University of Windsor, Ontario, Canada; "Sewage" in *ECT* 3rd ed., Vol. 24, pp. 407–419, by J. R. Pfafflin, consultant.

1. G. M. Fair, J. C. Geyer, and D. A. Okun, *Water and Wastewater Engineering*, Vols. 1 and 2, John Wiley & Sons, Inc., New York, 1968.
2. C. A. MacInnis, "Municipal Wastewater," in *The Encyclopedia of Environmental Science and Engineering*, Vol. 1, Gordon and Breach Science Publishers, New York, 1976, pp. 587–600.
3. *Standard Methods for the Examination of Water and Wastewater*, 15th ed., American Public Health Association, New York, 1981.
4. J. Salvato, *Environmental Sanitation*, 2nd ed., John Wiley & Sons, Inc., New York, 1972.
5. D. Thüry, *Resource Management and Optimization*, Vol. 1, No. 4, Harwood Academic Publishers, New York, 1981, pp. 289–314.
6. K. Imhoff, *Taschenbuch der Stadt Entwässerung*, R. Oldenburg Verlag, Munich, 1960.
7. N. Basta, *Chem. Eng.* 79 (June 14, 1982).
8. J. M. Lynch and F. Burde, *N.J. Effluents*, 14–15 (Aug. 1982).
9. *Wastewater Treatment Plant Design, Manual of Practice No. 8*, Water Pollution Control Federation, Washington, D.C., 1977.
10. T. Camp, *Trans. ASCE* **111**, 895 (1942).
11. J. Stevens, ed., *Hydrographic Data Book*, Lupold and Stevens, Inc., Portland, Ore., 1969.
12. K. C. Flynn, *J. Water Pollut. Control Fed.* **54**(4), 328 (1982).
13. K. Imhoff and G. Fair, *Sewage Treatment*, 2nd ed., John Wiley & Sons, Inc., New York, 1956.
14. E. B. Phelps, *Stream Sanitation*, John Wiley & Sons, Inc., New York, 1944.
15. J. C. Germain, *J. Water Pollut. Control. Fed.* **38**(2), 192 (1966).
16. R. E. McKinney, *Microbiology for Sanitary Engineers*, McGraw Hill Book Co., New York, 1962.
17. *Recommended Standards for Sewage Works*, Health Education Service, Albany, N.Y., 1977.
18. T. D. Reynolds, *Unit Operations and Processes in Environmental Engineering*, Wadsworth, Inc., Belmont, Calif., 1981.
19. J. K. Bewtra, "Biological Treatment of Wastewater," in *The Encyclopedia of Environmental Science and Engineering*, Vol. 1, Gordon and Breach Science Publishers, New York, 1976, pp. 109–126.
20. D. F. Othmer, *Catalyst*, Vol. 2, No. 1, Catalyst Publishing Co., New York, 1973, pp. 9–14.
21. J. R. Whirter, *The Use of High Purity Oxygen in the Activated Sludge Process*, Vol. 1, CRC Press, Inc., West Palm Beach, Fla., 1978.
22. *Code of Federal Regulations*, Title 40, Pts. 405–411.
23. W. W. Eckenfelder, Jr. in A. Bisio and S. Boots, eds., *Encyclopedia of Energy Technology and Environment*, Vol. 4, John Wiley & Sons, Inc., New York, 1995, pp. 2835–2837.
24. W. Goldfarb, "Federal Water Pollution Control Law Since 1972," in *Advances in Environmental Science and Engineering*, Vol. 2, Gordon and Breach

Science Publishers, New York, 1979, pp. 3–36.

25. W. F. Nolte, "Effects of Elevated Pressure on Secondary Treatment of Wastewater," in *Advances in Environmental Science and Engineering*, Vol. 4, Gordon and Breach Science Publishers, New York, 1981, pp. 161–169.

26. *Meeting the Challenge of the 80's*, U.S. Environmental Protection Agency, Washington, D.C., 1980.

27. C. Newkumet, *Sludge*, 16 (July Aug. 1980).

REUSE

An essentially constant supply of water is available for utilization by all living creatures; however, it is necessary to manage the water resources of a region in such a way that the available water meets all legitimate needs.

Use dictates the quality required. Potable water must be bacteriologically safe, and toxic substances must be present at levels that are accepted as safe (1–3) (see Table 1). In addition, the water must be aesthetically acceptable. Water that is suitable for drinking may be inadequate for many industrial processes. On the other hand, many industrial processes can use water that is not pure enough to drink.

Table 1. Maximum Allowable Concentrations of Significant Constituents in Municipal (Domestic) Water Supplies^a

Constituent	µg/L	Constituent	µg/L
arsenic	50	manganese	50
barium	1,000	mercury	2
cadmium	10	methoxychlor	100
chromium	50	nitrate (nitrogen)	10,000
copper	1,000	phenol	1
endrin	0.2	selenium	10
iron	300	silver	50
lead	50	toxaphene	5
lindane	4		

^a Ref. 2.

Water consumption is constantly increasing. In 1950, U.S. industry consumed $290 \times 10^6 \text{ m}^3 \text{ H}_2\text{O/d}$ ($7.7 \times 10^7 \text{ gal/d}$). By 1970, consumption had increased to $624 \times 10^6 \text{ m}^3 \text{ d}$ ($1.65 \times 10^8 \text{ gal d}$), and in 1980, to $964 \times 10^6 \text{ m}^3 \text{ d}$ ($2.55 \times 10^8 \text{ gal d}$) (4).

Formerly, water was accepted by a second user for reuse while it was still under control of the first user (5). Today, the used water is treated in such a manner that it can be used again before ultimate disposal. Furthermore, a distinction can be made between direct reuse, where the water is reclaimed without dilution or natural purification, and indirect use, where treated used water is returned to the environment for subsequent utilization as a raw water supply.

A notable example of controlled water reuse was utilization of secondary sewage effluent from the Back River Wastewater Treatment Plant in Baltimore by the Sparrows Point Works of Bethlehem Steel (6). The Sparrows Point plant was supplied primarily by wells located near the brackish waters of Baltimore harbor. Increased draft on the wells had led to saltwater intrusion. Water with chloride concentration as high as 10 mg/L is unsuitable for many steelmaking operations. Rollers, for example, are pitted by such waters. However, treated effluent from the Back River Plant can be used for some operations, such as coke quenching, and $>4 \times 10^6 \text{ m}^3 \text{ d}$ (10^8 gal d) are piped 13 km to Sparrows Point. This arrangement has proved economical to both parties for >40 yr.

Water can seldom be reused directly. The treatment required depends on the intended second use. Disposal costs of the wastewater must be included in any economic analysis, and additional treatment for reuse may be justified when this expense is included. Costs of reclamation depend on the location, water scarcity, availability of public water supplies, and the intended reuse.

Owing to increased public awareness of water pollution, stringent waste-disposal regulations have been introduced, and best available technology (BAT) had to be applied to industrial wastewater treatment by the mid-1980s (7). The cost–benefit analyses in BAT applications are controversial. Whereas three decades ago waste disposal was indicated only by an arrow in the process flow diagram, today it is a significant part of the operating costs. Pure water is expensive, and high quality sources are becoming more scarce. It is frequently necessary to accept lower quality water and subject it to costly treatment. Used water as a source of raw water supply has long been permitted by public health authorities in water-poor areas. However, in regions with an ample supply of fresh drinking water, there is resistance to the inclusion of reused water. A Gallup Poll indicated a 50% opposition to the use of reclaimed water in drinking water (8).

Conventional secondary wastewater treatment does not produce an effluent suitable for direct reuse, and a tertiary treatment step is required. Secondary treatment reduces BOD, COD (chemical oxygen demand), and >90% of suspended solids. Microbial concentrations are greatly reduced, and remaining organic material in the effluent is further decomposed by natural processes. Some nitrogen and phosphorus compounds are not removed and may cause eutrophication in the receiving waters. Contaminants that cannot be removed by conventional means are termed refractory, and they range from simple inorganic salts to complex organic substances such as pesticides, herbicides, and surfactants. Recycling of effluents from wastewater treatment plants in an essentially closed system normally would result in unacceptably high concentrations of many undesirable substances. The components of typical tap water are compared with those of a secondary effluent in Table 2 (9).

Table 2. Increase of Tap Water Components in Wastewater

Component	Tap water, mg/L	Secondary wastewater, mg/L	Difference, mg/L
BOD	2	20	18
COD	12	100	88
alkylbenzenesulfonates	0.4	6.8	6.4
cations			
Na ⁺	50	125	75
K ⁺	3	13	10
NH ₄ ⁺	0.1	16	15.9
Ca ²⁺	42	60	18
Mg ²⁺	16	19	3
anions			
Cl	66	143	77
NO ₃ ⁻	5	12	7
NO ₂	0.15	1.5	1.3

HCO ₃	198	296	98
CO ₃ ²	0.1	1	1
SO ₄ ²	56	84	28
SiO ₃ ²	29	43	14
PO ₄ ³ , total	8.1	25	17
PO ₄ ³ ^b	0.3	25	25
hardness ^c	158	235	77
alkalinity ^c	164	242	78
total dissolved solids	382	700	318
pH	8.0	7.4	0.6

^a Ref. 9.
^b Other than orthophosphates.
^c As CaCO₃.

Continued recycling of effluent would soon give an unusable product, and tertiary treatment, although expensive, is essential. In some instances, considerable amounts of water can be saved by careful housekeeping and process change.

In many cases, the quality of a stream or another water source can be adequately improved by removing more BOD or suspended solids. In other instances, the effluent is prepared for groundwater recharge which may require only the removal of nutrient. A classification of wastewater treatment processes is given in Table 3. Table 4 summarizes water quality criteria for various industrial uses (10).

Table 3. Wastewater Treatments^a

Process	Substance treated or removed
<i>Biological</i>	
conventional secondary treatment	suspended solids, BOD, bacteria
conventional process modifications	suspended solids, BOD, bacteria, nutrients
anaerobic denitrification	nitrates
algae harvesting	nitrates and phosphates
<i>Chemical</i>	
ammonia stripping	ammonia nitrogen
ion exchange	nutrients
electrodialysis	salts
chemical precipitation	suspended solids and phosphates
<i>Mechanical</i>	
activated carbon adsorption	organic matter and suspended solids
sedimentation	suspended solids and bacteria
filtration	suspended solids and bacteria
microstrainers	very small particles
reverse osmosis	salts
distillation	salts
foam separation	detergents
vapor-compression evaporation	concentration of wastewater
wasteheat evaporation	concentration of wastewater
steam stripping	hydrogen sulfide and ammonia

^a Ref. 7.

Table 4. Industrial Water Quality Limits^a

Industry	Turbidity, units	Colo r, units	Hardness	Tempera- ture, °C	pH	TDS	SS	SiO ₂	Fe	NaCl	Cl	SO ₄	Alkalinity
textiles (SIC 22)	0.3–5	0–5	0–50			100–200	0–5	25	0–0.3	0.01–0.05	100	100	50–200
pulp and paper (SIC 26)													
fine paper	10 ^b	5	100			200		20	0.1	0.03			75
kraft paper													
bleached	40 ^b	25	100			300		50	0.2	0.10	200		75
unbleached	100 ^b	100	200			500		100	1.0	0.50	200		150
groundwood papers	50 ^b	30	200			500		50	0.3	0.10	75		150
soda and sulfite paper	25 ^b	5	100			250		20	0.1	0.05	75		75
chemicals (SIC 28)		500	1,000	^c	5.5–9.0	2,500	10,000	^c	10	2.00	500	850	500
petroleum (SIC 2911)		25	900		6.0–9.0	3,500	5,000	85	15		1,600	900	500
iron and steel (SIC 33)				38	5.0–9.0		100 ^d						
food canning (SIC 2032, 2033)		5	250		6.5–8.5	500	10	50	0.2	0.20	250	250	250
tanning (SIC 3111)													
tanning processes	^e	5	150		6.0–8.0			50			250	250	^c
finishing processes	^e	5	^f		6.0–8.0			0.3	0.20		250	250	^c
coloring	^e	5	^e		6.0–8.0			0.1	0.001				^c
soft drinks (SIC 2086)		5	^g		^g	^g		0.3	0.05		500	500	85

^a From ref. 10; units are mg/L unless otherwise specified. Abbreviations: TDS, total dissolved solids; SS, suspended solids.
^b Units in mg/L as SiO₂.

^c Not considered a problem of concentrations encountered.
^d Settleable solids.
^e Not detectable by test.
^f Lime softened.
^g Controlled by treatment for other constituents.

Tertiary Treatment

Chlorination. Chlorination kills bacteria and is routinely included in secondary treatment. For some special uses, it gives a water quality that is acceptable for blending with other water in storage reservoirs.

Chemical Precipitation. When applied after conventional secondary treatment, chemical precipitation removes heavy metals and gives a high quality effluent. In this process, an insoluble compound is formed and the resulting precipitate allowed to settle. Charges on colloidal particles are neutralized, and particles agglomerate. The resulting floc sweeps other material, including bacteria, from suspension. Phosphates are removed by addition of soluble aluminum or iron salts. Insoluble aluminum phosphate or ferric phosphate are formed, but the removal may require adsorption on hydroxide floc. Addition of lime is the cheapest and most easily controlled method. At high pH, lime, in addition to removing hardness, forms a precipitate of hydroxyapatite, Ca₅(OH)(PO₄)₃. Lime also assists in removing sulfides and fluorides. The sludge is putrescible and a nuisance source, and disposal is difficult.

Filtration. Filtration is usually a misnomer for tertiary processes that remove particulate matter. Small particles are removed by adsorption rather than by physical straining. If secondary effluents contain a high concentration of solids, filter beds clog and binding occurs at the bed surface. Energy losses become high, and short circuiting passage of dirty water. Sand, mixed media, and diatomaceous earth are the most common filter materials. Rapid sand filters have a limited use in tertiary wastewater treatment. Suspended solids accumulate in the filter bed, and BOD in the water is reduced. Adsorbed foreign matter is removed by backwashing; frequency depends on the surface application rate and concentration of suspended material. For backwashing, water may be used alone at a rate of 100–120 mL (cm²·min) (25–30 gal (ft²·min)), or a combination of air and water may be used, with water applied at a rate of 40–60 mL (cm²·min) (10–15 gal (ft²·min)) and air at ca 120 mL (cm²·min) (30 gal (ft²·min)). Air serves to thoroughly agitate the filter. Particle range of the medium is 0.5–0.8 mm, bed depth is ca 60 cm, and the application rate of raw water is 8–10 mL (cm²·min) (2–2.5 gal (ft²·min)). Most of the particles are removed in the upper part of the filter, and bed porosity in this region decreases rapidly as the filtration progresses, excessive head loss results, and the filter clogs. Effective filter depth can be increased by mixed media, with the lightest and coarsest material at the top and the heaviest and least porous at the bottom. This is the opposite of the case where the bed contains material of constant density. In a mixed-media filter, three or more materials are used, including coal, sand, and garnet, with specific gravities of 1.65, and 2.65, and 4.00, respectively. Coal has the greatest porosity and garnet the least. After backwashing, where the filter acts as a fluidized bed, the filter materials stratify and the coal is on top. The resulting particles are graduated from ca 1 mm at the top to 0.15 mm at the bottom, and the entire filter bed is utilized for particulate removal. The garnet at the bottom forces the water to pass a much finer barrier than that provided by sand alone, whereas the coal layer does not clog as readily. Filter depths are 60–75 cm, and application rates of 4–120 mL (cm²·min)(1–30 gal (ft²·min)) are common.

In diatomaceous-earth filtration, the powdered filter aid is built upon a relatively loose septum to screen out suspended solids. The filter becomes clogged, and pressure losses become excessive; backwashing is then necessary. The smallest removable particle is 0.5–1 μm (see DIATOMITE).

Both vacuum and pressure filters are used. Turbidity is more easily removed by vacuum filters, usually at 85° efficiency. Flow rates are low, ca 4 mL (cm²·min) [1 gal (ft²·min)]; these filters are not practical for treating large volumes.

Microstrainers. Microstrainers are rotating steel screens with extremely fine stainless steel mesh (85–170 perforations per square centimeter (13–26 in.²)). The flowing liquid enters the open end of the drum and passes through the mesh to the effluent end. The mesh traps solid impurities and rotates with the drum. A wash-water spray washes the trapped solids into a hopper for final disposal. The mesh is washed with filtered effluent discharged from jets fitted into the drum and then exposed to uv radiation to inhibit microbial growth. The mesh is washed with chlorine water at intervals of 7 to 28 days in order to control slime growth; removal efficiencies are 30–55° of the applied BOD and 40–60° of suspended solids.

Effluent Polishing. The term polishing is sometimes applied to the preparation of effluents of exceptional clarity. Plant effluent is collected in a sump and pumped through a tube-settling unit, followed by mixed-media filtration. Filter effluent is collected in a storage tank, which also serves as a reservoir for backwash water. The tank further acts as a chlorine-contact chamber. The tube settler is a supplemental solids-separation device which allows the filter to operate efficiently during severe operational upsets, such as excessive loading of organic matter or increased hydraulic loading. In such instances, the bulk of the suspended solids in the plant effluent is removed in the tube settler and the filter is not overloaded. These installations produce an effluent low in BOD and suspended solids. The polishing units continue to function well when suspended solids are loaded at a rate of 2 g L (0.12 oz gal). The tube settler and mixed media can be used for phosphate removal after suitable treatment for coagulation and flocculation.

Foam Separation. Conventional secondary treatment cannot remove detergents that resist biological treatment such as alkylbenzenesulfonates. Although ABS has been almost completely replaced by linear alkylbenzenesulfonates in household detergents, foam (qv) continues to be a problem. Foam-separation techniques have been applied for removing refractory matter such as organic hydrates and nitrogen compounds (11). A sparger disperses small bubbles of gas, usually air, throughout the waste. The rising gas collects suspended solids and surface-active substances while the foam collects at the liquid–liquid interface. It is removed and collapsed to yield a waste concentrate.

For small installations, column foam separators are more suitable. Waste flows downward in the column whereas gas spargers, located at the bottom, give countercurrent flow. The foam generated is carried upward to a foam breaker and collector.

Through-type separators are more effective for larger municipal and industrial treatment facilities. The operating principles are the same, but flow is through a trough rather than a column.

Activated-Carbon Adsorption. Treatment of waste by activated-carbon adsorption has shown promise of becoming an important tertiary treatment process. It has long been applied for the removal of tastes, odor, and color but has not been suitable for treatment of effluents containing high concentrations of organic matter because of rapid decrease of surface area. As the surface of the carbon becomes coated with adsorbed material, less surface area is available for further adsorption. Activated carbon is available either as fine powder or granules. Use of the latter is preferred for wastewaters. In a packed-bed installation, water is passed through a column filled with activated-carbon particles, and the organic content of the water is reduced along the path of flow. However, the carbon at the influent end becomes exhausted and must be regenerated. A fluidized bed gives more uniform concentration gradient throughout the bed. Packed beds accumulate solids more rapidly than fluidized beds. Activated-carbon adsorption can remove 80° of the COD and 70° of BOD. Regeneration of carbon is accomplished in multistage hearth furnaces in a steam–air mixture at 76°C. Approximately 5° of the carbon is lost in regeneration.

Ion Exchange. Ion exchange (qv) is a method of softening hard water. Calcium and magnesium are exchanged for a cation, usually sodium. Ion-exchange beds can remove 95° of phosphate, 85° of nitrates, 100° of sulfates, and 45° of COD. Color and organic matter are efficiently removed by cationic and anionic mixed beds, but the latter tend to foul the beds. Cost and frequency of regeneration are principal disadvantages of ion exchange.

Oxidation Ponds. Oxidation ponds, or sewage lagoons, long popular in Europe, now also find wide application in North America. Careful design is necessary for successful operation. Although oxidation ponds are, in essence, large open ponds into which waste is led, careful planning allows natural purification processes to proceed in an orderly manner giving a stable and high quality effluent. The main disadvantage is production of odors offensive to nearby residents. In addition, relatively large areas are required.

Many facilities utilize air diffusers or surface aerators for supplying additional oxygen. The capacity of an oxidation pond is mostly a function of the oxygen transferred from the atmosphere to the pond water but also depends on other physical and biological variables. These ponds, usually designed for 2.2–3.3 gBOD (m²·d) (6.5–9.7 oz (acred)), are ca 1.2 m deep. Mixing is very important. If waste material were uniformly distributed throughout the pond,

the oxygen demand would be constant throughout the pond. This is, however, not the case, and oxygen requirements vary from area to area. The processes involved in waste stabilization are mostly aerobic, but facultative organisms are dominant at the bottom of the pond. Most of the oxygen is transferred from the surface, but additional oxygen may be added by aerators and diffusers; most surface aerators operate intermittently (12). Natural surface agitation is important, and prevailing wind direction must be considered for surface aeration as well as deflection of odors affecting downwind residents. The pond must be deep enough to avoid weed problems but shallow enough to allow sunlight to penetrate. Algae maintain aerobic conditions. The pond functions as a sedimentation basin, and it may be necessary to periodically remove deposited solids. Current practice is to feed influent at the pond center and withdraw effluent at the periphery. Multiple ponds are used in series (7,8).

Water discharged from an oxidation pond may be used directly for low quality industrial purposes or mixed with other waters (13).
Reverse Osmosis. In reverse osmosis (qv), a solution or suspension flows under pressure through a membrane; the product is withdrawn on the other side. This process can treat dissolved solids concentrations ranging from 1 mg L to 35 g L (14). The principal constraint is the requirement that the waste material be relatively nonfouling. Recent advances have been mostly in membrane development, and pilot studies are required (15). Energy costs can be significant, and it is frequently necessary to pretreat influent in order to minimize fouling. Reverse osmosis can deal with particles < 1 to 600 nm in size.

Electrodialysis. In reverse osmosis pressure achieves the mass transfer. In electrodialysis (qv), dc is applied to a series of alternating cationic and anionic membranes. Anions pass through the anion-permeable membranes but are prevented from migrating by the cationic permeable membranes. Only ionic species are separated by this method, whereas reverse osmosis can deal with nonionic species. The advantages and disadvantages of reverse osmosis are shared by electrodialysis.

Vapor-Compression Evaporation and Waste Heat Evaporation. Both of these processes remove water from contaminants rather than contaminants from water. They are better suited for industrial installations where excess energy is available. The water thus produced is of high quality and can be used directly. An important advantage is the concentration of waste-residue volume with attendant economies of handling and transportation (see EVAPORATION).

The saturation temperature of a vapor rises when it is mechanically compressed and its latent heat is available at a higher temperature. Application of this heat to an aqueous stream evaporates part of the water, producing a distillate of pure water. Application of vapor compression has grown significantly since 1960.

Cooling loads can be transferred from process heat exchangers to a wastewater-evaporation system, thus reducing cooling-water requirements and total wastewater volume.

In both processes, care must be taken in dealing with organic compounds that either steam distill or form azeotropes.

BIBLIOGRAPHY

"Water Reuse" in *ECT* 2nd ed., Vol. 22, pp. 116–124, by James R. Pfafflin, University of Windsor, Ontario, Canada; "Water, Reuse" in *ECT* 3rd ed., Vol. 24, pp. 420–426, by James R. Pfafflin, Consultant.

1. *Quality Criteria for Water*, Environmental Protection Agency, Washington, D.C., 1976.
2. *Fed. Regist.* **4**, 57332 (Aug. 27, 1980).
3. *Guidelines for Water Reuse*, EPA 600 8-80-036, Environmental Protection Agency, Washington, D.C., 1980.
4. W. Hales, *Chemtech*, 532 (Sept. 1982).
5. R. Kasperperson, *Public Acceptance of Water Reuse*, University Press of New England, Hanover, N.H., 1977.
6. A. Wolman, *Sewage Works J.* **20**(1), 1 (1948).
7. A. D. Holiday, *Chem. Eng.*, 118 (Apr. 19, 1982).
8. *J. Am. Water Works Assoc.* **65**(8), 513 (1973).
9. *Studies of Wastewater Reclamation and Utilization*, pub. no. 9, State Water Pollution Control Board, Sacramento, Calif., 1954.
10. W. W. Eckenfelder, Jr., in A. Bisio and S. Boots, eds., *Encyclopedia of Energy Technology and the Environment*, Vol. 4, John Wiley & Sons, Inc., New York, 1995, p. 2826.
11. R. B. Grieves, C. J. Randall, and R. K. Wood, *Air Water Pollut. Int. J.* **8**(8–9), 501 (1964).
12. T. D. Reynolds, *Unit Operations and Processes in Environmental Engineering*, Wadsworth, Inc., Belmont, Calif., 1982.
13. W. W. Eckenfelder, Jr., *Principles of Water Quality Management*, CBI Publishing, Boston, Mass., 1980.
14. S. Sourirajan, "Reverse Osmosis—Process Fundamentals" in *The Encyclopedia of Environmental Science and Engineering*, Gordon and Breach Science Publishers, Inc., New York, 1976, p. 738.
15. E. C. Kaup, *Chem. Eng.*, 46 (Apr. 2, 1973).

General Reference

Water: Conservation and Reclamation, The Global Cities Project, San Francisco, Calif., 1991.

TREATMENT OF SWIMMING POOLS, SPAS, AND HOT TUBS

The swimming pool market has grown at an average of about 3% per year over the last 10 years. The total number of pools in 1991 was about 6×10^6 in the United States (1). Residential pools represent about 90% of the total and are about equally divided between above-ground and in-ground types, but the latter are growing at a faster rate. The majority of inground pools are in Florida and California. Hundreds of cities operate public pools, and many schools, clubs, and camps have outdoor and indoor pools. Most motels and many hotels, condominiums, and apartment complexes also have pools on their premises. Although commercial pools account for only about 10% of the total, they represent about half of the treated water. The number of spas and hot tubs has grown rapidly since 1970, with an estimated 2.7×10^6 spas and hot tubs in 1991 in the United States and the market is growing. About 17% of spas hot tubs are attached to pools. California has the greatest number of spas and hot tubs.

Swimming Pools

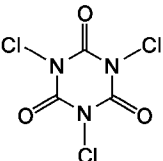
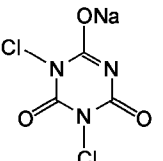
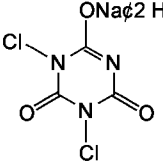
Most swimming pools are of the recirculating type. Through filtration, chemical treatment, and dilution with rain and makeup water, the water can be reused without draining and refilling. The only time that pools need to be drained and refilled is for repair or in cases of contamination, severe scaling or algae infestation. Some health codes require partial replacement of pool water (based on bather load) to maintain water quality in public swimming pools. Sanitizing chemicals must be added regularly to oxidize pool contaminants and kill and control disease-carrying bacteria and other organisms introduced by swimmers and dirt entering the water. It is also necessary to destroy algae whose spores are carried into the water by wind and rain. Unchecked algal growth results in discolored water, unsightly growth on pool surfaces, clogging of filters, and provides breeding ground for bacteria. The pH of pool water must be maintained within a desirable range for swimmer comfort and optimal effectiveness of chlorine sanitizers. In order to control the corrosive or scaling tendencies of pool water, it is also necessary to maintain a proper balance between pH, alkalinity, and hardness. Undesirable trace metals such as iron, manganese, or copper are sometimes found in source water or formed by corrosion of pool equipment. Unless removed, these metals discolor pool water and cause stains, especially damaging to plaster pool surfaces. Filtration of pool water is necessary for removal of suspended solids which otherwise cloud water and interfere with the disinfection process. The National Swimming Pool Foundation trains personnel as operators for public and private pools and spas through the Certified Pool Spa Operator program (2).

Construction. New residential inground pool construction consists of concrete gunitite 63%, vinyl liner 33%, and fiberglass and other 4% (1). The pools vary in size from $<4.6 \times 9.1$ m ($<15 \times 30$ ft) to $>6.1 \times 12.2$ m ($>20 \times 40$ ft) and have an average volume of about 109.8 m³ (29,000 gal). Aboveground pools vary in diameter from <3.7 m (<12 ft) to >7.3 m (>24 ft), have an average volume of 56.8 m³ (15,000 gal), and typically have vinyl liners. The average size of commercial pools is 363.4 m³ (96,000 gal).

Design. The basic recirculation loop of a swimming pool consists of strainers (for removing hair and lint), skimmers (in outdoor pools for removing large floating debris such as leaves, twigs, and insects), a pump, a filter, and in some cases, a heater (typically gas). The National Sanitation Foundation sets standards and provides testing for circulation system components for pools, spas, and hot tubs (3). Inground pools can have bottom drains allowing removal of some debris that settles to the bottom of the pool. Large pools require multiple inlets especially in the deep end to ensure adequate circulation and elimination of dead spots. Large municipal pools typically have balancing tanks to hold water displaced by bathers and are equipped with instrumentation for automatic measurement and control of available chlorine and pH via sanitizer and chemical feeders.

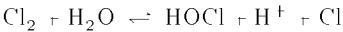
Chlorine-Based Sanitizers. Chlorine compounds are the most commonly used swimming-pool sanitizers and include: chlorine gas, sodium hypochlorite solution, calcium hypochlorite, trichloroisocyanuric acid (Trichlor), and sodium dichloroisocyanurate (Dichlor) (Table 1) (see also DISINFECTANTS). Nonchlorine sanitizers are used to a small extent. Sanitizer usage varies geographically (4), eg, in the Sun Belt overall consumption varies as follows: chloroisocyanurates > calcium hypochlorite > sodium hypochlorite, whereas, in the Snow Belt consumption generally varies in the order: calcium hypochlorite > chloroisocyanurates > sodium hypochlorite. However, in the Midwest, chloroisocyanurates predominate. Calcium hypochlorite usage is highest on the East Coast and lowest on the West Coast. Chloroisocyanurate users also use significant amounts of hypochlorites (typically calcium hypochlorite) for superchlorination or shock treatment since chloroisocyanurates are not recommended for this purpose.

Table 1. Chlorine Sanitizers Used in Swimming Pools

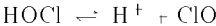
Compound	Formula	Typical form	Cl, % av
chlorine	Cl ₂	liquefied gas	100
trichloroisocyanuric acid		tablets, sticks	89–91
calcium hypochlorite	Ca(OCl) ₂	granules, tablets, briquettes	65–75
sodium dichloroisocyanurate		granules	62–63
sodium dichloroisocyanurate dihydrate	 ONa•2 H ₂ O	granules	55–56
lithium hypochlorite	LiOCl	granules	35
sodium hypochlorite	NaOCl	solution	10–15

Chlorine. Chlorine gas is used as a sanitizer mainly in large commercial pools requiring a high feed rate to maintain the desired residual. However, many pool service companies use chlorine for treatment of residential pools. Some state and local health codes specify chlorine gas as the

required or preferred disinfectant for public pools above a certain size. When chlorine dissolves in water, it hydrolyzes to produce hypochlorous acid, hypochlorite ion, and hydrochloric acid (5):



$$K_{\text{eq}} = 3.94 \times 10^{-4} \text{ at } 25^\circ\text{C}$$



$$K_{\text{eq}} = 2.88 \times 10^{-8} \text{ at } 25^\circ\text{C}$$

The dissociation of hypochlorous acid depends upon pH and, to a much lesser extent, temperature (6). At 25°C, it is ~0% at pH 5, about 50% at pH 7.5, and ~100% at pH 10, see Figure 1. Because of the acidity formed by chlorine gas, addition of soda ash (Na₂CO₃) or sodium sesquicarbonate (Na₂CO₃·NaHCO₃) is necessary to maintain the proper pH and to replenish alkalinity.

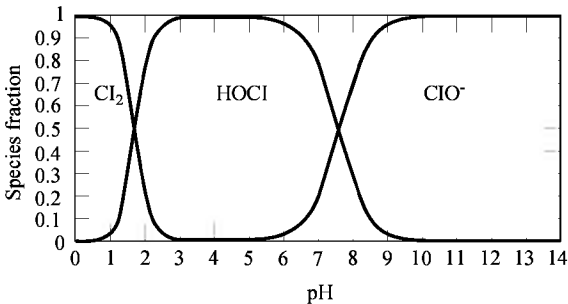
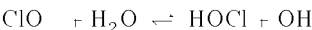
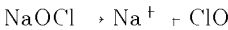


Fig. 1. Distribution of Cl₂, HOCl, and ClO⁻ vs pH.

Chlorine is supplied as a liquefied gas in steel cylinders of 45–68 kg capacity. It requires special metering and feeding equipment and experienced supervision. Because of its toxicity, elaborate precautions against leakage are needed. Although Cl₂ gas is not expensive, initial equipment costs, cylinder demurrage, and potential hazards make this automatic disinfection system impractical for all but very large, heavily used pools. By contrast, some pool service companies treating residential pools simply use a small portable chlorine cylinder of ~9 kg net wt with flexible tubing and a chlorine diffuser. However, this method of pool treatment is risky because of the low pH region generated during rapid chlorine dissolution which can extend to the pool walls and bottom resulting in etching of plaster or tile grout pools surfaces or bleaching of vinyl liners. Calcium and sodium hypochlorite are finding increased use in municipal and public pools because of the health and safety hazards of chlorine gas (see ALKALI AND CHLORINE PRODUCTS). Small electrolytic chlorine generators are available and are based on electrolysis of brine (ie, aqueous NaCl). The gaseous chlorine formed in the anode compartment is dissolved in the recirculating pool water and the sodium hydroxide formed in the cathode compartment can be used for adjusting the pH of the pool.

Sodium Hypochlorite Solution. Sodium hypochlorite (liquid bleach) is usually bottled as a 10–15% available Cl₂ solution, containing 0.65–0.83% NaOH for stability which has only a small effect on pool pH. It is a commonly used sanitizer for swimming pools. In pool water, it produces hypochlorite ion and hypochlorous acid:



Sodium hypochlorite is widely used in California and Florida, the two states with the largest number of in-ground pools, where it is a popular sanitizer with the home-pool owner because of low cost per unit package. It is also widely used for small public pools and for semicommercial pools such as at motels and camps. Disadvantages of sodium hypochlorite solution are low storage stability and bulkiness which requires a large storage area near large pools. Small single-chamber electrochemical generators available for producing NaOCl (from electrogeneration of chlorine gas and sodium hydroxide) are based on electrolysis of pool water containing ~2000 ppm NaCl.

Calcium Hypochlorite. This chemical, marketed since 1928, is one of the most widely used swimming-pool water sanitizers. Calcium hypochlorite, a crystalline solid, is a convenient source of available chlorine and is sold in granular or tablet form for use in home, semiprivate, and commercial pools. When dissolved in water, Ca(OCl)₂ forms hypochlorous acid and hypochlorite ion similar to NaOCl. It contains small amounts of stabilizing Ca(OH)₂, which has a very small effect on pool pH (7). Calcium hypochlorite has superior storage stability and much higher available Cl₂ concentration than liquid bleach, which reduces storage requirements and purchasing frequency.

Lithium Hypochlorite. Similar in action to other hypochlorites, this chemical is used to a small extent. It is more expensive than sodium and calcium hypochlorite.

Chlorinated Isocyanurates. The cyanuric acid-based sanitizers, introduced for pool use in 1958, are stable crystalline compounds with moderate-to-high available Cl₂. Sodium dichloroisocyanurate (Dichlor), sold in granular form, dissolves rapidly, whereas trichloroisocyanuric acid (Trichlor) dissolves very slowly and is widely used in the form of tablets or sticks in feeders, floating devices, or in the pool skimmer.

Other Sanitizers and Treatment Systems. Other sanitizers and swimming pool treatment systems used to a limited extent are bromine, quats, ozone, ionizers, electrolyzers, uv, uv H₂O₂, and Ag H₂O₂.

Bromine. Bromine for swimming pool disinfection is supplied by 1-bromo-3-chloro-5,5-dimethylhydantoin or is generated *in situ* from bromide ion and an oxidizing agent. Use of liquid bromine or electrochemical hypobromite generation is not significant. A disadvantage of bromine based sanitizers in outdoor pools and spas is that bromine (ie, HOBr and BrO⁻) cannot be stabilized against decomposition by sunlight. Bromochlorodimethylhydantoin, BCDMH, is sold in tablet form and has an av. Cl of 29% and an av. Br of 66%. Hydrolysis provides hypobromous acid for disinfection. The N–Cl bond hydrolyzes to a much smaller extent than the N–Br bond, forming HOCl which reacts rapidly with residual bromide ion in the water forming more HOBr (HOCl + Br⁻ → HOBr + Cl⁻). Because of low solubility, BCDMH requires soaking-type feeders. Hypochlorite sanitizers are used for shock treatment. The hydantoin ring can be cleaved by hypochlorite ion yielding N-chloroisopropylamine, NCl₃, and CO₂ (8). A bromine generating system used to a small extent involves dosing the pool with sodium bromide combined with routine addition of potassium peroxymonosulfate to generate HOBr (HSO₅⁻ + Br⁻ → BrO⁻ + HSO₄⁻, BrO⁻ + H₂O → HOBr + OH⁻). Other oxidizing agents such as hypochlorites, chloroisocyanurates, and ozone can also be used.

Quats. The polymeric quaternary poly(hexamethylenebiguanide) hydrochloride, is a bactericide that requires routine addition of an algicide. Since chlorine sanitizers cannot be used for shock treatment, H₂O₂ is used as the oxidant. However, H₂O₂ is a poor oxidizing agent for ammonium ion and organic matter whose concentrations can increase with time. Some organic matter can reduce disinfection efficiency. In addition, the quaternary ammonium algicide used with this system can cause foaming and is removed by filter media. Other problems with this system have been reported including off taste, staining, cloudiness, and biological growths (eg, slime mold and pink algae).

Ozone. Ozone generators are based on uv or silent discharge. Although uv ozone generators are marketed, they are not effective for treating

swimming pools because of the low ozone concentration produced (40–400 ppm in air). By contrast, silent discharge ozone generators producing 1–2 wt % ozone in air are effective and are used extensively in Europe, primarily in large public or commercial pools (see OZONE). Ozone cannot be used as a primary disinfectant because of its instability, volatility, and toxicity. Thus it is used primarily for oxidation of swimming pool contaminants. Ozone is added to the water in the external recycle loop and after about a 2-minute reaction time in a contactor, it is destroyed by filtration through granular activated carbon (9). Chlorine is then added to the water for disinfection. In smaller public and semipublic pools, ozone is used in combination with NaBr to generate hypobromite ion, which hydrolyzes to hypobromous acid for disinfection ($\text{O}_3 + \text{Br}^- \rightarrow \text{BrO}^- + \text{O}_2$, $\text{BrO}^- + \text{H}_2\text{O} \rightarrow \text{HOBr} + \text{OH}^-$). By contrast with European standards, U.S. practice varies according to the type of installation (ie, new or retrofit) and can employ variations such as prefilter ozone injection combined with single mixed-media filtration and post filter ozone injection and side stream ozonation.

Ionizers. Ionizers generate small concentrations of copper and silver ions (by electrochemical dissolution of a copper–silver electrode) that function as algicide and bactericide, respectively (10). Although the concentration of copper (~0.3 ppm) is adequate, the concentration of silver, a poor disinfectant, is very low (<50 ppb). Consequently chlorine sanitizers are necessary not only for effective disinfection but also for oxidation of swimming pool contaminants. Another disadvantage is that copper and silver ions form colored insoluble precipitates, which can cause staining.

Electrolyzers. Electrolytic units have been marketed that claim to generate active species such as hydroxyl radicals and oxygen atoms. However, these systems have been shown to be ineffective in either disinfection, algae prevention, or oxidation of swimming-pool contaminants.

Uv, Uv-H₂O₂, and Ag/H₂O₂. The disadvantages of these systems are (1) uv and Ag/H₂O₂ provide little to no oxidation of pool contaminants; (2) bacteria are capable of repairing damage by uv light, reducing the inactivation rate; (3) both uv and uv/H₂O₂ do not supply a disinfectant residual in the pool, providing only localized disinfection whose rate is affected by turbidity and pool contaminants; (4) both silver and H₂O₂ are poor disinfectants; and (5) precipitation of silver on pool surfaces can cause staining.

Sanitizer and Chemical Feeders. Feeders dispense the chemicals in gaseous, liquid, and solid (both granular and compacted) forms. Many health departments require that public pools have approved feeding devices for the daily application of all chemicals, including sanitizers. A slurry feeder for diatomaceous earth (DE) on diatomite filter installations may also be required.

Gas Feeders. Chlorine gas is usually fed from a chlorine cylinder equipped with a pressure gauge, reducing valve, regulating valve, feed-rate indicator, and aspirator-type injector for dissolving the chlorine gas in water. The feeder can be manually, or more desirably automatically, controlled utilizing continuous amperometric or potentiometric measurement of the free chlorine residual. The chlorine solution is normally introduced into the return line to the filter.

Liquid Feeders. Liquid feeders employ positive-displacement metering pumps for adding aqueous solutions of sodium or calcium hypochlorite. The feed solutions are typically stored in polyethylene tanks of various capacities up to about 0.19 m³ (50 gal).

Solid Feeders. Feeders are available for feeding solid chemicals such as soda ash, diatomaceous earth (DE), and sanitizers. Erosion-type feeders installed in the pool return line are commonly used for solid sanitizers in compacted form; typically tablets of calcium hypochlorite or trichloroisocyanuric acid. Off-line installation is used when it is impractical to install the feeder in the pool return line. Some feeders have an integral body with a chamber for holding tablets (11), whereas others employ a removable cartridge filled with tablets. The dissolving rate of the sanitizer is controlled by the depth of immersion of the tablets and varying the flow rate of water by means of a valve. A feeder for Ca(OCl)₂ tablets or briquettes, used extensively in commercial pools, utilizes pulse feeding, ie, a slow water fill into the tablet dissolving chamber followed by a fairly rapid siphoning of the resultant concentrated solution into the pool return line. The feed rate is controlled by the depth of immersion of the tablets and the filling–draining frequency. Scaling is prevented by using an antiscalant (12). Another type of feeding technique involves addition of tablets or sticks of sanitizing agent (usually trichloroisocyanuric acid) to the skimmer basket located in the filter return line. If calcium hypochlorite tablets are added to the skimmer, the filter pump must remain on until complete dissolution to avoid excessively high local concentrations of available chlorine which could cause corrosion. Slow dissolving cylindrical tablets of calcium hypochlorite (~200 g) for use in the skimmer are encased in a plastic wrap containing a small opening at either end for contact with water (13). Another skimmer device for continuous feeding of Ca(OCl)₂ uses tablets or briquettes in a plastic cylindrical container with a variable opening in the top. A different type of feeder consists of a cartridge containing tablets or sticks, which is inserted to a variable depth into a hollow plastic flotation ring. Several units may be used which float freely around the pool continuously dispensing sanitizer. Feeders that dispense granular sanitizing agents such as calcium hypochlorite are also available.

pH Control Feeders. Feeders for controlling pH when using chlorine gas sanitation are typically based on pumping a solution of soda ash. Dry chemical feeders can also be used, eg, adding granular soda ash from a hopper by means of a reciprocating screw feeder to a solution chamber fed by pool water with the overflow mixing with the recirculating pool water. Liquid feeders for controlling pH when using hypochlorites typically pump muriatic acid. Gaseous carbon dioxide can also be used for pH adjustment.

Water Quality Maintenance. In addition to controlling algae and microorganisms such as bacteria, proper swimming pool maintenance requires control of free and combined available chlorine, pH, alkalinity, hardness, and saturation index. Ranges for various swimming pool parameters (Table 2) are recommended by The National Spa and Pool Institute (14).

Table 2. Recommended Swimming Pool, Spa, and Hot Tub Parameters^a

Parameter	Minimum	Ideal	Maximum
free chlorine, ppm			
pools	1.0	1.0–3.0	3.0
spas hot tubs	2.0	3.0–5.0	10.0
combined chlorine, ppm	0	0	0.2
bromine, ppm			
pools	2.0	2.0–4.0	4.0
spas hot tubs	2.0	3.0–5.0	10.0
pH	7.2	7.4–7.6	7.8
total alkalinity, ppm	60	80–100 ^b	180
total dissolved solids, ppm	300	1000–2000	3000
calcium hardness, ppm	150	200–400	500–1000 ^c
cyanuric acid, ppm	10	30–50	150 ^c

^a Ref. 14.
^b For hypochlorite sanitizers; 100–120 ppm for acidic sanitizers, chlorine, Dichlor, Trichlor, and bromine compounds.
^c Except where limited by Health Dept. requirements, often to 100 ppm.

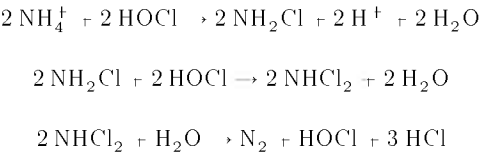
Disinfection. The bactericidal properties of chlorine (15), bromine (16), and ozone (15) have been reviewed. The biocidal properties of chlorine and bromine are due principally to the formation of hypochlorous and hypobromous acids which kill bacteria, algae, and other microorganisms. Disinfection with bromine is less sensitive to pH than with chlorine. Although, bromamines are better disinfectants than chloramines, they are less stable. The rate of disinfection with chlorine and bromine is primarily a function of the free available halogen concentration, the type of organism (ie, bacteria, viruses, protozoa, etc), contact time, pH, and temperature. The presence of interfering substances such as turbidity (ie, suspended solids), ammonia, and organic matter can adversely affect disinfection (17). Whereas halogens kill bacteria by diffusion of hypohalous acid through the cell membrane, ozone inactivates bacteria by rupturing the cell wall (ie, lysing). The efficiency of disinfection is determined bacteriologically by the standard plate-count (SPC) technique and by examination for fecal coliform bacteria (18).

Maintenance Available Chlorine. All chlorine sanitizers provide free available chlorine (FAC), ie, HOCl + ClO⁻. Available chlorine is a

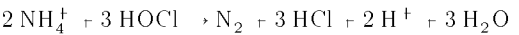
measure of the oxidizing power of the electropositive chlorine in a compound expressed in terms of elemental chlorine. The recommended ideal free available chlorine concentration is 1.0 ppm in an unstabilized pool and 1–3 ppm in a stabilized pool (14). Free available chlorine and combined available chlorine (CAC, ie, chloramines) in swimming-pool water are determined colorimetrically by the Palin DPD test using *N,N*-diethyl-*p*-phenylenediamine (18). The av. Cl can be maintained by liquid bleach, periodic broadcasting of granular sanitizers (typically calcium hypochlorite) or by means of tableted sanitizers (calcium hypochlorite or trichloroisocyanuric acid) placed in the skimmer, in floaters, or in-line or off-line feeders.

Superchlorination–Shock Treatment. Superchlorination or shock treatment of pool water is necessary since accumulation of organic matter, nitrogen compounds, and algae consumes free available chlorine and impedes disinfection. Reaction of chlorine with constituents of urine or perspiration (primarily NH₄⁺, amino acids, creatinine, uric acid, etc) produces chloramines (N–Cl compounds) which are poor disinfectants because they do not hydrolyze significantly to HOCl (19). For example, monochloramine (NH₂Cl) is only 1/280 as effective as HOCl against *E. coli* (20).

During superchlorination or shock treatment, ammonium ion is oxidized to nitrogen by breakpoint chlorination which is represented by the simplified reaction sequence



Overall



In reality the chemistry of breakpoint chlorination is much more complex and has been modeled by computer (21). Conversion of NH₄⁺ to monochloramine is rapid and causes an essentially linear increase in CAC with chlorine dosage. Further addition of chlorine results in formation of unstable dichloramine which decomposes to N₂ thereby causing a reduction in CAC (22). At breakpoint, the process is essentially complete, and further addition of chlorine causes an equivalent linear increase in free available chlorine. Small concentrations of combined chlorine remaining beyond breakpoint are due primarily to organic chloramines. Breakpoint occurs slightly above the theoretical Cl:N ratio (1.75 vs 1.5) because of competitive oxidation of NH₄⁺ to nitrate ion. Organic matter consumes chlorine and its oxidation also increases the breakpoint chlorine demand. Cyanuric acid does not interfere with breakpoint chlorination (23).

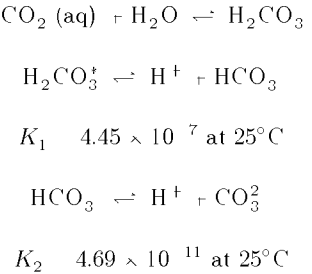
Urea (24), amino acids (25), and creatinine (26) are also decomposed during superchlorination or shock treatment, with formation of N₂ and other oxidation products. However, the process is slower than with ammonium ion (see CHLORAMINES AND BROMAMINES). Urea is the principal nitrogen-containing compound in swimming pools. Since it is an amide, it reacts slowly with chlorine, yielding N₂, NCl₃, and NO₃[−] (27).

Superchlorination typically refers to adding FAC equal to 10× ppm CAC, whereas shock treatment generally involves addition of 10 ppm FAC. The frequency of superchlorination or shock treatment depends on bather load and temperature. Calcium hypochlorite, because of its convenience, is widely used for superchlorination and shock treatment. Sodium hypochlorite, LiOCl, or chlorine gas are also used. Chloroisocyanurates are not recommended since their use would result in excessive cyanuric acid concentrations.

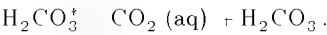
Potassium peroxymonosulfate, introduced in the late 1980s, is finding increasing use as an auxiliary oxidant for shock treatment and oxidation of chloramines. Sodium peroxydisulfate is also being sold for shock treatment, however, it is less reactive than peroxymonosulfate. Mixtures of sodium peroxydisulfate and calcium hypochlorite can be used for shock treatment (28). Disadvantages of peroxymonosulfate and peroxydisulfate are they do not provide a disinfectant residual and peroxymonosulfate oxidizes urea and chloramines to nitrate ion, which is a nutrient for algae.

pH Control. The optimal range for bather comfort and efficiency of disinfection of chlorine sanitizers is pH 7.2–7.8 where the biocidal agent HOCl represents 69–35% of FAC. This percentage increases at lower pHs, but increased eye irritation and corrosion may occur. At higher pH, the HOCl fraction decreases (eg, to 26% at pH 8.0) and the ClO[−] fraction correspondingly increases leading to greater photochemical decomposition and reduced rate of disinfection. Hypochlorite ion is only 1/80 as effective as HOCl against *E. coli* (17).

The pH of swimming pool water is a function of the following equilibria:



where



The concentration of true carbonic acid (H₂CO₃) is negligible in comparison to dissolved carbon dioxide, eg, only 0.3% of the latter is hydrated to carbonic acid at 25°C. The ionization constant K₁ is a composite constant representing both the CO₂ hydration reaction, K_h and ionization of true H₂CO₃, K_{H₂CO₃}, ie, K₁ = K_{H₂CO₃} (1 + K). Temperature-dependent equations for K₁ and K₂ are (29)

$$\begin{aligned} \log K_1 &= 356.3094 - 0.060920T + 21834.37/T \\ &\quad + 126.8339 \log T - 1684915/T^2 \\ \log K_2 &= 107.8871 - 0.032528T + 5151.79/T \\ &\quad + 38.92561 \log T - 563713.9/T^2 \end{aligned}$$

where T is in kelvins. The total concentration of species is given by:

$$C_T = [\text{H}_2\text{CO}_3^*] + [\text{HCO}_3^-] + [\text{CO}_3^{2-}]$$

The concentration of individual species is represented by:

$$\begin{aligned} [\text{H}_2\text{CO}_3^*] &= C_T \alpha_0 \\ [\text{HCO}_3^-] &= C_T \alpha_1 \\ [\text{CO}_3^{2-}] &= C_T \alpha_2 \end{aligned}$$

where α_0 , and α_1 are the ionization fractions, ie, the mol fraction of $[H_2CO_3^*]$, $[HCO_3^-]$, and $[CO_3^{2-}]$, respectively. The ionization fractions can be expressed in terms of the hydrogen ion concentration (calculable from the pH) and applicable equilibrium constants.

$$\alpha_0 = \frac{1}{1 + K_1 [H^+] + K_1 K_2 [H^+]^2}$$
$$\alpha_1 = \frac{[H^+]}{[H^+] + K_1 + K_1 K_2 [H^+]}$$
$$\alpha_2 = \frac{[H^+]^2}{[H^+]^2 + K_1 K_2 + [H^+] K_2}$$

A plot of ionization fractions for $[H_2CO_3^*]$, $[HCO_3^-]$, and $[CO_3^{2-}]$ as a function of pH at 25°C is shown in Figure 2.

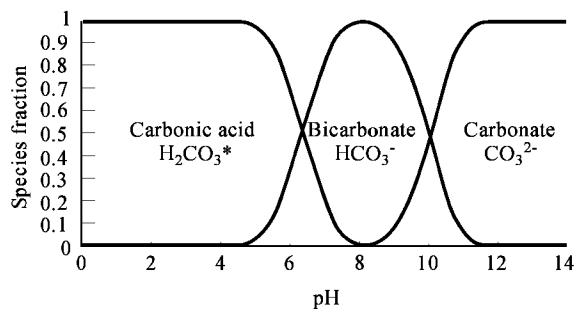
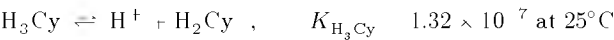


Fig. 2. Distribution of carbonic acid, bicarbonate, and carbonate vs pH.

In stabilized pools, the following equilibrium is also important (30):



Although cyanuric acid is a tribasic acid, only the first ionization is important at normal swimming pool pHs. The concentration of cyanurate is given by $[H_2Cy^-] = C_T' \alpha_1'$, where C_T' is the total concentration of cyanuric acid and cyanurate ion and α_1' is the mol fraction of cyanurate ion which is given by $\alpha_1' = ([H^+] + K_{H_3Cy})^{-1}$, where K_{H_3Cy} is the first ionization constant of cyanuric acid whose temperature dependence is expressed by (31):

$$\log K_{H_3Cy} = -31.08 + 0.154T - 2.441 \times 10^{-4} T^2$$

where T is in kelvins. A plot of ionization fractions for the cyanuric acid system at 25°C is shown in Fig. 3.

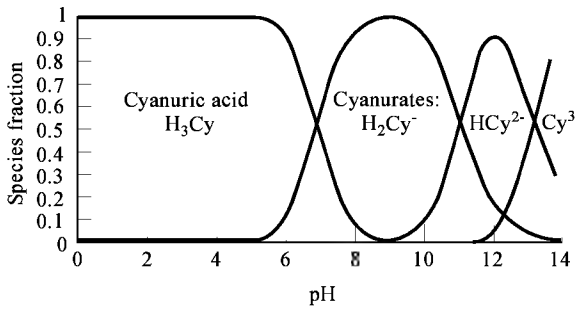


Fig. 3. Distribution of cyanuric acid and cyanurates vs pH.

Total alkalinity (Alk_t) is expressed by:

$$Alk_t = [HCO_3^-] + 2[CO_3^{2-}] + [H_2Cy^-] + [OH^-] - [H^+]$$
$$= C_T(\alpha_1 + 2\alpha_2) + C_T' \alpha_1' + [OH^-] - [H^+]$$

At typical pool pH, the concentration of hydroxyl and hydrogen ions can be neglected. In addition to cyanurate, other alkaline substances (eg, borate, phosphate, etc) can also contribute to total alkalinity (32). Alkalinity provides buffering, ie, resistance to pH change. The buffer intensity (β) is the incremental change in alkalinity required to change pH by one unit, ie, $\beta = \Delta Alk / \Delta pH$. For stabilized swimming pool water, it is expressed numerically by:

$$\beta = 2.3 \{ C_T [\alpha_1 (\alpha_0 + \alpha_2) + 4\alpha_0 \alpha_2] + C_T' \alpha_0' \alpha_1' + [OH^-] + [H^+] \}$$
$$\approx 50 \times 10^3 \text{ ppm } CaCO_3$$

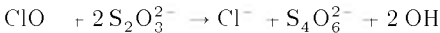
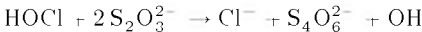
where α_0' is the mol fraction of cyanuric acid. The buffer intensity of pool water decreases with increasing pH.

Although swimming pools are open to the atmosphere they are not in equilibrium with atmospheric carbon dioxide (33). Because of periodic acid addition for pH adjustment, swimming pools are maintained in a supersaturated state with respect to dissolved carbon dioxide. Thus, swimming pools continuously evolve carbon dioxide causing the pH to increase with time, the rate increasing with increasing alkalinity, turbulence, and temperature (34). However, this upward pH drift is modified according to the acidity or basicity of the sanitizer. Hypochlorites (Ca, Li, and Na) raise pH to a small extent, whereas acidic sanitizers lower it, ie, $Cl_2 > Trichlor > Dichlor$ (7). In the case of chlorine gas (or in cases of heavy Trichlor usage) the effect of CO_2 loss can be completely offset causing the pH to drift downward.

The pH is measured colorimetrically with phenol red indicator. High FAC causes lower pH readings due to bleaching of the indicator and resultant HCl formation. The pH of pool water is readily controlled with inexpensive chemicals. Hydrochloric acid solution or sodium bisulfate lower it, whereas sodium carbonate raises it. Since acid addition neutralizes a portion of the alkalinity, this must be replenished if the alkalinity drops below the minimum. By contrast, pH adjustment with carbon dioxide does not affect alkalinity.

Alkalinity. In swimming-pool water at its normal pH range, the so-called carbonate alkalinity is due primarily to bicarbonate with a very small contribution from carbonate. Alkalinity is expressed in terms of ppm of equivalent $CaCO_3$. Because of its buffering intensity, alkalinity resists changes in pH when sanitizing chemicals are added to pool water. Since cyanuric acid is largely neutralized at pool pH, cyanurate ion also contributes to alkalinity (and buffering). Therefore, in stabilized pools the alkalinity determination is corrected for $1.3 \times \text{ppm}$ cyanuric acid in the normal pH range (7.2–7.8) (32).

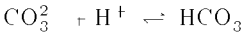
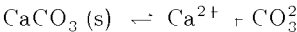
Sodium bicarbonate is generally added to increase alkalinity and muriatic acid (HCl) or sodium bisulfate (NaHSO₄) to reduce it. In general, with acidic sanitizers such as chlorine gas or trichloroisocyanuric acid, ideal total alkalinity should be in the 100–120 ppm range, whereas, with alkaline products such as calcium, lithium, or sodium hypochlorite, a lower ideal total alkalinity of 80–100 ppm is recommended (14). Alkalinity is determined by titration with standard sulfuric acid using a mixed bromcresol green–methyl red indicator after dechlorination of the sample with thiosulfate. Dechlorination with thiosulfate causes higher readings due to formation of hydroxyl ion (32):



The error is approximately equal to the FAC at pH 7.2–7.8. However, it is significant only at high FAC, eg, during superchlorination or shock treatment. **Hardness.** Water hardness is caused by certain polyvalent metals and is expressed in terms of ppm of equivalent CaCO₃. In swimming-pool water, the hardness is caused primarily by calcium ions and to a lesser extent magnesium ions. To ensure proper water balance the calcium hardness should be at or near the CaCO₃ saturation value. At 25°C, 7.5 pH, 1000 ppm total dissolved solids, 100 ppm total alkalinity, 50 ppm cyanuric acid, the calculated saturation Ca hardness is ~301 ppm which is within the NSPI recommended ideal range of 200–400. Values at other pool conditions can be calculated using the saturation index formula. Calcium hardness is determined by titration with ethylenediaminetetraacetic acid (EDTA) at pH 12–13 using Eriochrome Blue Black R, Eriochrome Black T, calmagite, hydroxynaphthol blue, or other suitable indicator. Hardness is raised with CaCl₂ and is lowered by draining some of the pool water and adding water of lower hardness.

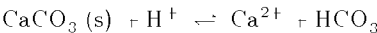
Saturation Index. Materials of construction used in pools are subject to the corrosive effects of water, eg, iron and copper equipment can corrode whereas concrete and plaster can undergo dissolution, ie, etching. The corrosion rate of metallic surfaces has been shown to be a function of the concentrations of Cl⁻, SO₄²⁻, dissolved O₂, alkalinity, and Ca hardness as well as buffer intensity, time, and the calcium carbonate saturation index (35). Deposition of a protective layer of crystalline CaCO₃ has been proposed for protection of metallic surfaces against corrosion by using the natural calcium and alkalinity in water (36).

A big concern in swimming pools is prevention of etching and scaling (ie, precipitation of CaCO₃) which can be controlled by maintenance of a proper degree of saturation with respect to calcium carbonate. The calcium carbonate dissolution-precipitation equilibrium is represented by:



$$K_{\text{eq}} = K_1 \cdot K_2$$

Overall



$$K_{\text{eq}} = K_s \cdot K_2$$

The temperature dependent equation for the solubility product constant *K_s* is (29):

$$\log K_s = 171.9065 - 0.077993T + 2839.319/T + 71.595 \log T$$

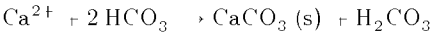
where *T* is the temperature in kelvins. The tendency of a water to precipitate or dissolve calcium carbonate can be determined by the calcium carbonate saturation index (SI). The saturation index is the common logarithm of the degree of CaCO₃ saturation, ie, $[\text{Ca}^{2+}][\text{CO}_3^{2-}]/K_s = [\text{Ca}^{2+}][\text{HCO}_3^-]/K_2[\text{H}^+]/K_s$, where the terms in brackets represent molar concentrations. A revised and updated version of the saturation index is obtained by substitution of [Hard] (calcium hardness) and [Alk] (alkalinity) in ppm CaCO₃, for [Ca²⁺] and [HCO₃⁻], respectively, (37):

$$SI = \text{pH} + \log[\text{Alk}] + \log[\text{Hard}] + TC + C$$

Where *C* = *A* + *B* - 9.7, *A* = log(*K*₂ / *K_s*), *K*₂ is the second ionization constant of H₂CO₃, *TC* = 0.02 + 0.01485*T_c* is a temperature correction term that adjusts log(*K*₂ / *K_s*) from its value at 0°C (- 2.25) to higher temperatures *T_c* (°C), *B* = log γ_{HCO₃⁻} + log γ_{Ca²⁺} / γ_{HCO₃⁻} and γ_{Ca²⁺} are the activity coefficients of bicarbonate and calcium ions calculated using the Davies approximation, log γ = - 0.5*z*²[√μ / (1 + √μ) - 0.3μ], *z* is the ionic charge, and μ is the ionic strength. The ionic strength can be calculated by μ = 0.5Σ*c_iz_i*² = 2.5 × 10⁻⁵ × *TDS* = 1.6 × 10⁻⁵ × κ, where *c_i* is the concentration of an individual ion and *z_i* its ionic charge, *TDS* is ppm total dissolved solids, and κ is the conductivity in micro Siemens /cm. At 1000 ppm TDS, *C* = 12.29. SI is also equal to pH_a - pH_s, where pH_a is the measured pH, and pH_s is the pH that the water should have to be at CaCO₃ saturation (ie, water that does not deposit or dissolve CaCO₃); pH_s = pH_a - SI. Ion pair formation typically has a small effect on SI.

The saturation index is rooted in thermodynamics (38). It is a scaling index and not a corrosion index. The numerical value of SI for a given water determines its departure from equilibrium, ie, > 0 = oversaturated, 0 = saturated, and < 0 = undersaturated. The higher the positive value of the index (SI > 0), the greater the tendency for CaCO₃ precipitation or scaling and the lower the negative value (SI < 0), the greater the tendency for etching. Water with a sufficiently negative index will not only etch concrete, plaster, and tile grout but it can also dissolve an existing protective film of CaCO₃.

Clear water with moderate to high positive values of SI does not undergo spontaneous homogeneous precipitation because the nucleation rate is too low (39). However, precipitation can occur in the presence of a sufficient concentration of suspended solids (ie, turbidity). The more these suspended solids resemble calcium carbonate in crystal structure the more likely the precipitation. When CaCO₃ precipitation occurs, the extent and rate is a function of the saturation index, the presence of impurities, turbidity, and the buffer intensity (40) which in turn depends on the alkalinity and pH of the water. Impurities such as Mg ions, polyphosphates (present in some source waters), and naturally occurring organic matter can inhibit precipitation of calcium carbonate. Heterogeneous precipitation (ie, scaling) may occur on some rough pool surfaces. Elevated temperature in pool heaters may cause precipitation because the solubility of CaCO₃ decreases with increasing temperature:



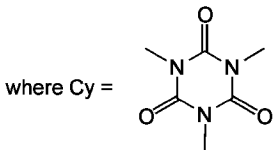
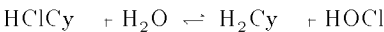
Excessive scaling can lead to flow restriction and reduced heat transfer.

To avoid problems such as etching or scaling, pool water should have a small positive saturation index. A practical operating range is 0 to 0.5, which corresponds to 100 to ~300% of saturation. Water can be brought to a state of approximate equilibrium (balance) by suitable adjustments in pH, alkalinity, and hardness. However, the SI will drift, primarily because the pH changes due to CO₂ loss and addition of sanitizer (7). An upward drift typically occurs when hypochlorites or a combination of hypochlorites and chloroisocyanurates are used for pool sanitation. By contrast, a downward drift occurs with chlorine gas or in cases of high Trichlor usage. A practical approach is to adjust the SI to the midpoint of the recommended range and to monitor the pH to establish the actual drift pattern. In order to maintain proper water balance for a maximum percent of the time, the SI should be adjusted to the beginning of the normal drift range. Many swimming-pool chemical dealers and some pool service companies have computer programs for calculating necessary water balance adjustments. When a pool is winterized, the saturation index must be adjusted for the lower anticipated temperatures.

Ancillary Chemicals. In addition to sanitizers, various ancillary chemicals are employed in swimming pools for stabilizing chlorine sanitizers, water balance, control of algae, scale and stain prevention, water clarification, filter degreasing, and cleaning of pool surfaces. Compatibility with chlorine must be considered in selecting ancillary chemicals since many of these additives are organic materials which have chlorine demands and therefore may interfere with pool disinfection. For example, quats and chelating agents such as citrate and phosphonate are oxidized by chlorine sanitizers, the oxidation being accelerated by sunlight.

Stabilizers. Cyanuric acid is used to stabilize available chlorine derived from chlorine gas, hypochlorites or chloroisocyanurates against decomposition by sunlight. Cyanuric acid and its chlorinated derivatives form a complex ionic and hydrolytic equilibrium system consisting of ten isocyanurate species. The 12 isocyanurate equilibrium constants have been determined by potentiometric and spectrophotometric techniques (30). Other measurements of two of the equilibrium constants important in swimming-pool water report significantly different and or less precise results than the above study (41–43). A critical review of these measurements is given in Reference 44.

A plot of species distribution for the cyanuric acid-available chlorine system at 25°C is shown in Figure 4. In the presence of excess cyanuric acid, the predominant chlorinated specie is the monochloroisocyanurate ion, HClCy where Cy represents the triisocyanurate anion. Therefore, the only significant equilibria in pool water are



At 25°C, pH 7.5, 1.5 ppm FAC, and 25 ppm cyanuric acid, the calculated HOCl concentration is only 0.01 ppm. Although the monochloroisocyanurate ion hydrolyzes to only a small extent, it serves as a reservoir of HOCl because of rapid hydrolysis. Indeed, this reaction is so fast that HClCy behaves like FAC in all wet methods of analysis. Furthermore, since HClCy absorbs uv only below 250 nm, which is filtered out of solar radiation by the earth's atmosphere, it is more resistant to decomposition than the photoactive ClO, which absorbs sunlight at 250–350 nm and represents the principal mode of chlorine loss in unstabilized pools (30). As little as 5 ppm of bromide ion prevents stabilization of FAC by cyanuric acid (23) (see also CYANURIC AND ISOCYANURIC ACIDS).

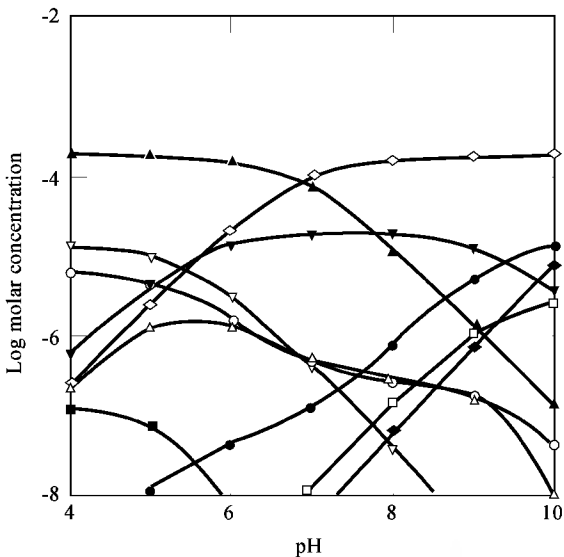


Fig. 4. Species distribution in the cyanuric acid (CA)–av Cl system at 25°C, CA ~25 ppm, av Cl ~1 ppm. ○ = HOCl, ● = ClO⁻, ▽ = H₂ClCy, ▼ = HClCy, □ = Cl₂Cy²⁻, ■ = HCl₂Cy, △ = Cl₃Cy, ▲ = H₃Cy, ◇ = H₂Cy⁻, and ◆ = HCl₂Cy⁻.

Based on the above equilibria, the concentration of HOCl in the normal pH range varies inversely with the total concentration of cyanurate. Increased concentration of cyanuric acid, therefore, should decrease the biocidal effectiveness of FAC. This has been confirmed by laboratory studies in buffered distilled water which showed 99% kill times of *S. faecalis* at 20°C increasing linearly with increasing cyanuric acid concentration at constant av. Cl at pH 7 and 9 (45). Other studies in distilled water have found a similar effect of cyanuric acid on kill times of bacteria (46–48). Calculations based on the data from Ref. 45 show that the kill times are highly correlated to the HOCl concentration and poorly to the concentration of the various chloroisocyanurates, indicating that HOCl is the active bactericide in stabilized pools (49).

By contrast to distilled water, studies in swimming pool water showed greater kill times, even in the absence of cyanuric acid (47). This was attributed to unknown variables, which in reality are swimming pool contaminants. In swimming pools, nitrogenous compounds such as ammonium ion, creatinine, and amino acids, introduced into the pools by bathers, can react with free av. Cl. This forms bactericidally ineffective combined chlorine, resulting in increased kill times. Studies have shown that addition of ammonia greatly increases the kill time of *S. faecalis* by chlorine whether cyanuric acid is present or not (48). By contrast urea, a principal pool contaminant, had no effect. Although urea is a potential source of ammonia, its hydrolysis is very slow under swimming-pool conditions. However, since urea is a nutrient for bacteria and algae, it is necessary to oxidize it by periodic shock treatment.

The ideal recommended cyanuric acid concentration is 30–50 ppm (Table 2). Although this range can be readily maintained when using hypochlorite sanitizers, it cannot be maintained when using chloroisocyanurates since they increase the cyanuric acid concentration. The NSPI recommends a maximum of 150 ppm cyanuric acid. Many health departments limit cyanuric acid to 100 ppm. No significant increase in stabilization occurs beyond 50–100 ppm, and since high levels of cyanuric acid slow down the rate of disinfection, the pool water should be partially drained and replaced with fresh water to reduce the cyanuric acid to below recommended maximum levels. Cyanuric acid is determined turbidimetrically after precipitation as melamine cyanurate.

Water Balance Chemicals. Water balance chemicals include muriatic acid, sodium bisulfate, and soda ash for pH control, sodium bicarbonate for alkalinity adjustment, and calcium chloride for hardness adjustment. A recent development is use of buffering agents for pH control. One of these products, sodium tetraborate, hydrolyzes to boric acid and a small amount of orthoborate (50) which provides significantly less buffering than carbonate and cyanurate alkalinity in the recommended pool pH range of 7.2–7.8 even at 100 ppm.

Algicides. Algal growth in pools is unsightly, a potential safety hazard to swimmers, and usually a result of poor pool maintenance. It can cause slipperiness, development of odors, cloudy and discolored water, chloramine formation, increased chlorine demand, bacterial growth, and stubborn stains. Low FAC, high temperatures, sunlight, and certain mineral nutrients promote algae growth. Such growth can be prevented by maintaining the proper pH range, free available chlorine content, and cyanuric acid concentration supplemented by periodic shock treatment.

The three forms of algae commonly found in swimming pools are the fast-growing green algae, slow-growing blue-green algae (sometimes called black algae), and mustard or yellow algae. Green algae are relatively easy to control because they remain suspended in the water. By contrast, black and mustard algae are more resistant to treatment because they grow on the pool walls and bottoms where they become firmly attached, especially on rough surfaces such as plaster or tile grouting and can penetrate cracks. As with bacteria, studies have shown that algae (particularly mustard algae) are more difficult to control in stabilized pools. If algae become entrenched, shock treatment of the water with a hypochlorite (repeated as necessary) in conjunction with mechanical treatment (ie, brushing) remove most algae. Patches of algae on pool walls are sometimes treated by trichloroisocyanuric acid contained in a porous bag suspended in the affected area. Two products for treatment of algae infestations contain sodium bromide and ammonium sulfate. Addition of sodium bromide to a chlorine sanitized pool is not recommended because it will cancel the stabilization of chlorine.

Algicides other than chlorine (or bromine) are used to a small extent, principally in the hotter southern and western regions of the U.S. They serve as a backup to chlorine primarily as a preventive or corrective measure against unbalanced pool conditions. The most widely used algicides are the quaternary ammonium compounds (quats), of which the *n*-alkyldimethylbenzylammonium chloride type (with C₁₂–C₁ alkyl groups) are the most common (see QUATERNARY AMMONIUM COMPOUNDS). Formulation with other algicides improves their effectiveness, eg, *n*-alkyl-dimethyldichlorobenzylammonium chloride and copper bis[2,2',2''-nitritotris(ethanol)-*N,O,O'*] (copper triethanolamine complex) (51). Quats are absorbed on filter media (52–53) and consequently require high initial doses and frequent replenishments. The quats are surfactants and therefore can cause foaming at sufficiently high concentrations. A polymeric quaternary, poly(oxyethylene(dimethylimino)ethylene(dimethylimino)ethylene dichloride), is also employed and is nonfoaming at use concentrations. Other algicides used to a smaller extent contain silver compounds (eg, AgNO₃ or colloidal silver), copper compounds (ie, CuSO₄ and copper chelates with citric acid or triethanolamine), and La₂(CO₃)₃ (54). Zinc compounds can also be used as algicides (55). The use of Simazine, 2-chloro-4,6-bis(ethylamino)-*s*-triazine, has been discontinued since the EPA has removed its registration because of toxicity concerns. The effectiveness of various chemicals against algae has been evaluated in numerous studies under laboratory conditions (51–53,56–58). The concentrations necessary to kill algae (algicidal doses) are generally appreciably higher than those required to prevent or control the growth of algae (algistatic doses) and vary from specie to specie. Some algicides have chlorine demands which must be taken into account in maintaining the proper FAC (53).

Scale and Stain Controllers. Polyacrylates (low molecular weight) and organic phosphonates, eg, (1-hydroxyethylidene)diphosphonic acid, prevent or control precipitation of CaCO₃ by acting as chelating agents (qv) or dispersants (qv) to prevent excessive formation of hard scale by promoting crystal distortion.

Clarifiers. Pool water may occasionally contain metallic impurities such as copper, iron, or manganese which enter the pool with the makeup water or by corrosion of metallic parts in the circulation system. These dissolved metals can discolor the water and cause stains. Chlorine oxidizes soluble Fe²⁺ and Mn²⁺ to the highly insoluble Fe(OH)₃ and MnO₂ which can be removed by filtration. Water-soluble, high molecular weight polymers can be used as coagulating agents for removal of precipitated trace metals or other undesirable turbidity. These polymers can also serve as flocculating agents (qv) to increase the settling rate or strength of a chemical floc, thereby acting as a filter aid to control the depth of floc penetration (59). The polymers may be anionic, cationic, or nonionic, eg, polyacrylates, poly(styrene sulfonate), poly(diallyldimethylammonium chloride), poly(ethylene oxide), etc (60). Inorganic compounds such as aluminum sulfate and polyaluminum chloride are sometimes also used as flocculating agents for clarification or improving filtration.

Filter Cleaners. Grease and oils from bathers can affect filtration. Degreasers are employed to clean diatomaceous earth (DE) and sand filters. They can be surfactant or enzyme based.

Tile and Vinyl Cleaners. Cleaners are employed to remove the ring of soil that forms at the water line on tile and vinyl lined pools.

Test Kits. Proper pool management requires routine analysis for free and combined chlorine and pH, and less frequently, alkalinity, hardness, and cyanuric acid. These analyses can conveniently be carried out at poolside with simple test kits which are available at moderate cost. The more elaborate kits with photometers for color intensity measurement and meters for pH measurement are more accurate but more expensive. Test strips for chlorine and pH determinations are not as accurate as tests employing solutions. Water clarity (ie, turbidity) can be measured by test kit or visually. Digital titrators are available for alkalinity and hardness which read directly in ppm. Test kits are also available for N and Fe determinations. Other analyses such as total dissolved solids, trace metals (eg, Mn, Cu, etc) are carried out as needed by analytical laboratories. Bacteriological analysis of pool water can be obtained through local health departments or commercial laboratories. Kits are also available for determining bacteria levels. Detailed procedures for chemical, bacteriological, and biological analysis of water are given in Ref. 18.

Filtration. Efficient filtration for removal of suspended particles is essential for sparkling pool water. Filters may be of the fixed-bed, precoat, or cartridge type and operate under vacuum or pressure. The most common filter media are sand, anthracite, diatomaceous earth (DE), and paper or cloth cartridges. The cartridges remove particles as small as 15 μm, sand to 25 μm, and DE to 3 μm (61). Since the visual limit is 30–35 μm, most properly operating filters do a satisfactory job. Sand and DE filters are backwashed, whereas cartridge filters are cleaned manually by rinsing in water. Backwashing removes accumulated insoluble matter which would eventually increase resistance to flow (head) and reduce filtration rates. Flocculents for retention of surface dirt on sand filters are not recommended (2). Diatomaceous earth filters require a precoat of DE after each filter cycle (see DIATOMITE). Use of compressed air (bumping) to dislodge the DE followed by recoating the filter with the dislodged DE reduces water usage and allows longer filter cycles. The DE is sometimes mixed continuously with unfiltered water before entering the filter. The pump serves not only to circulate pool water through the filter and chlorinator but also to mix the pool water, thereby diluting and dispersing sanitizers and other additives whether added manually or through feeders. Typically, pumps are sized to achieve one pool water turnover in about 6 h. Vacuum cleaners, which derive their suction from the filter pump, are used to remove solids from the walls and bottom of the pool. Some automatic pool cleaners moving around the bottom of the pool are efficient enough to eliminate the need for hand vacuuming.

Spas and Hot Tubs

The basic principles of swimming-pool water treatment also apply to spas and hot tubs. However, spas and tubs are not miniature swimming pools but are unique in treatment requirements because of use patterns and a high ratio of bather to water. For example, four people in a 1.9-m³ (500-gal) spa or tub have a sanitizer demand equal to 160 people in a 75.7-m³ (20,000-gal) swimming pool.

Construction. Spas are manufactured in various shapes and sizes and are generally constructed of molded fiberglass resin or fiberglass acrylic. Larger in-ground spas are typically made of gunnite or concrete with a marbelite plaster finish. An average spa has a depth of 0.74 m, a diameter of 2.13 m, and a water capacity of 2.08 m³ (550 gal). Typical hot tubs are circular, 1.52–1.83 m in diameter with a 1.22 m depth and are constructed of redwood, oak, pine, cedar, cypress, teak, or mahogany. Wooden spa tubs are lined with vinyl or acrylic. The hot-water support systems of spas and hot tubs are basically the same, consisting of a pump, filter, heating system, and air blower. Pumps are sized to achieve one water turnover in about 30 min. Available filter systems include high rate sand, diatomaceous earth, and cartridge types. The heated water is violently mixed by an integrated forced-air system and venturi-type jets. The heating systems can be electric or gas.

Sanitizers. Spa and hot-tub sanitation is dominated by chlorine- and bromine-based disinfectants. Public spas and tubs usually employ automatic feeders, eg, Cl₂ gas feeders, to maintain a disinfectant residual. Private or residential spas and tubs can use automatic chemical feeding or generating devices, or they can be sanitized manually with granular or liquid products. The most widely used products for private spa and tub sanitation are sodium dichloroisocyanurate and bromochlorodimethylhydantoin. Granular products are normally added before and after use, whereas solids, eg, stick-bromine, are placed in skimmers or feeders. Bromine generating systems can also be used and are based on oxidation of bromide ions (added to the water as sodium bromide) by peroxymonosulfate, chloroisocyanurates, hypochlorites, or ozone to generate the disinfectant HOBr.

Water Quality Maintenance. The NSPI recommends an ideal residual of 3–5 ppm FAC for chlorine sanitizers with CAC not exceeding 0.2 ppm; see Table 2 (14). For bromine sanitizers, the maintenance of a 3–5 ppm ideal residual is recommended.

Statistical analysis of spa data showed that the rate of FAC loss was simply a linear function of the number of bathers during use periods; cyanuric

acid stabilization did not affect the loss of chlorine. However, the use of chloroisocyanurates for both sanitation and shock treatment resulted in a buildup of cyanuric acid to above the 150 ppm NSPI limit within 15 user periods (ie, four bathers with 15 min of exposure at 40°C), necessitating frequent changing of the water. Furthermore, these studies showed combined chlorine levels to range from 0.8 to 1.6 ppm after a four-bather exposure exceeding the NSPI limit of 0.2 ppm CAC. Shock treatment with ten times the CAC content after each user period was found satisfactory in reducing these CAC levels to <0.2 ppm. Based on these studies, it was concluded that a spa or tub should have a 4–5 ppm FAC before bather entry and maintaining the FAC in the recommended range while the spa or tub is in use (62). The spa or tub water should be shock treated after each use.

In addition to replacing spa or tub water because of cyanuric acid build-up from chloroisocyanurate sanitizers, the NSPI recommends that the water be replaced at least monthly or more frequently when often used because bathers contribute microorganisms, perspiration, body oils, lotions, etc, which affect water chemistry, total dissolved-solids concentration, water surface tension (resulting in foaming), and sanitizer demand. High bather density causes these compounds to accumulate to relatively high concentrations compared to a swimming pool that has a lower bather-to-water ratio. The guideline for replacing spa hot tub water is based on the formula: $n = 88V/b$, where n is the number of days before replacement, V is the volume in m³, and b is the number of bathers/day (3).

In wooden tubs, the maintenance of a sanitizer residual becomes complicated due to the leaching of tannins and other organic matter from the wood into the water. The sanitizer demand of these substances must be overcome in order to maintain proper residual concentrations. As the tub ages, the leaching of these materials decreases, but bleaching of the wood may occur as the lignin (qv) in the wood reacts with sanitizers.

pH control in spas and hot tubs is complicated because the violent turbulence and aeration which removes CO₂ from the water causing the pH to increase. The CO₂ loss rate also increases with temperature and carbonate alkalinity. The buffering intensity of cyanuric acid reduces the rate of pH rise. Maintaining the alkalinity at the lower end of the recommended range will moderate the upward pH drift.

Control of Microorganisms. Treatment-regime tests with Ca(OCl)₂ and chloroisocyanurates killed all pathogenic bacteria, including the *Pseudomonas* that cause skin infections, within 30 s by 2 ppm FAC at pH 7.4 (62). In a series of tests under stabilized conditions (50–100 ppm cyanuric acid), an initial concentration of 3 ppm FAC could not provide water that met the NSPI standards for bacteria or the maintenance of at least a 2 ppm FAC residual. Increasing the initial FAC to 4 ppm resulted in the bacterial criteria being met 100% of the time even when the FAC residual fell below 1 ppm. For heavy bather loads an initial FAC > 4 ppm is recommended.

Ancillary Chemicals. Various chemicals are used in spa and hot tub maintenance including aqueous silicone solutions (polydimethylsiloxane) for defoaming, sequestrants (organic phosphonates) for scale and stain prevention, flocculating agents (polyelectrolytes) for clarification, cleaners, tints, and fragrances. Since these chemicals have significant chlorine demands, they might reduce the bacterial kill efficiency by lowering the FAC or by interfering in the disinfection process. The sanitizer demand of these chemicals must be compensated for to ensure the maintenance of the residual concentrations stated on the product's label.

Economic Aspects

U.S. sanitizer consumption for swimming pools, spas, and hot tubs is given in Table 3 (4). Spas and hot tubs account for ~4% of the total. The market is growing at an average of 2–3% yr. Estimated nonchlorine oxidizer consumption in 1995 is sodium peroxydisulfate and potassium peroxymonosulfate ~500 t and hydrogen peroxide ~1600 t (63). The consumption of algicides is estimated to be ~4000 t. Quats account for about 78% of the market which is growing at about 5% yr.

Table 3. 1994 U.S. Sanitizer Consumption in Swimming Pools, Spas, and Hot Tubs

Compound	10 ³ t
chloroisocyanurates	48.8 ^a
calcium hypochlorite	50.0
sodium hypochlorite (Cl ₂ equivalent)	37.3
chlorine gas	13.6 ^b
lithium hypochlorite	2.9
bromochlorodimethylhydantoin	7.7 ^c

^a 1992.
^b 1995 (63).
^c 1993.

Health Factors

Improperly maintained swimming pools, spas and hot tubs can result in a variety of illnesses ranging from minor to life threatening.

Disease Transmission. Reported illnesses and outbreaks have been attributed to bacteria (*Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Legionella*, and *Shigella*), protozoa (*Giardia* and *Cryptosporidium*), and viruses (adenovirus, enterovirus, and Hepatitis A virus) (64,65). *Giardia* and *Cryptosporidium* are more difficult to inactivate than enteric bacteria such as *E. coli*. One of the most frequently isolated organisms found most often in spas, is *Pseudomonas aeruginosa*. *P. aeruginosa* is a hardy thermophilic bacteria and is most often responsible for outbreaks of folliculitis, skin dermatitis, otitis, pneumonia, and urinary tract infections. The principal source of biological contamination entering the water is bathers who introduce microorganisms through their skin, body fluids, and fecal contamination. Since resistance to chlorine varies with the organism, it is necessary to maintain the sanitizer level above the minimum recommended level. Pools and spas with high bather loads require automatic chlorinators and continuous monitoring of disinfectant residual through measurement of the oxidation-reduction potential. In addition monitoring for *E. coli* is necessary as an indicator of fecal contamination which poses the greatest health risk.

Spa/Hot Tub Microbiology. Spa or hot-tub water is an excellent environment for the growth of microbes because of optimal temperatures and the introduction of nutrients from bathers. Diseases can be transmitted by contact with and ingestion of the water, inhalation of microorganisms contained in aerosols produced by the aerated water, and infection through close personal contact. Analysis of unsanitized spa water has shown that a load of 10³–10⁴ microorganisms per mL in a 1.1-m³ (300-gal) spa can result from a 15-min use by a single bather. Currently, there are no national standards governing sanitation of spas and hot tubs. Standards proposed by the National Spa and Pool Institute (NSPI) are similar to those for swimming pools (14). Public spas and tubs are under the jurisdiction of local health departments. From the microbiological standpoint, the ideal recommendations exclude bacteria and visible algae. However, local health codes may allow a maximum limit of 200 bacteria per mL as determined by the SPC procedure, and no positive results in a most probable number (MPN) determination for coliforms, ie, organisms that indicate fecal contamination (18).

Spa/Hot Tub Usage. The use pattern of a spa or tub is dictated by personal preference. The water temperature of spas and tubs is normally maintained at 36.7–40.0°C. Although information from spa dealers suggests 15 min at 40°C, and 30 min at 36.7°C, individuals are known to spend longer periods. Long exposure at these temperatures may result in nausea, dizziness, or fainting. Spas and tubs should never be used while under the influence of alcohol (66). Elderly persons and those suffering from heart disease, diabetes, or high or low blood pressure should not use a spa or hot tube without prior consultation with a physician (66). Most important, the water temperature should not exceed 40°C.

Skin Irritation in Spas/Hot Tubs. A common skin irritation contacted in spas and hot tubs is a nonpruritic rash due to *P. aeruginosa* (67–70). Some cases of skin irritation have also been associated with use of bromine sanitizers. Skin infections may be enhanced by the high water

temperatures which open pores and remove protective oils.

Eye and Respiratory Tract Irritation. Swimming pool water is a foreign environment to the eyes, and physiological changes can occur that result in irritation. In a comprehensive study, the irritation was about 30° less at pH 8 than at 7 (71). Within the pH range 7–8, pH had a greater effect on eye irritation than free available chlorine (<0.5 ppm) in unstabilized water. There is no evidence that stabilized chlorine is irritating to the eyes. However, inorganic chloramines can irritate the eyes or mucous membranes and cause objectionable odors; the effect varying in the following order: $\text{NCl}_3 > \text{NHCl}_2 > \text{NH}_2\text{Cl}$. Chloramines (primarily NHCl_2 and NCl_3) are usually responsible for complaints of eye irritation. Swimmers may blame this condition on too much chlorine, but the problem is caused by insufficient chlorine. Because inorganic chloramines are decomposed by sunlight, they pose less of a problem for bathers in outdoor swimming pools than in indoor pools.

Disinfection By-Products. Disinfection by-products with chlorine and bromine sanitizers include chlorate and nitrate ions, inorganic and organic halamines and halogenated organic compounds (primarily trihalomethanes, THMs). Use of ozone can lead to additional by-products, eg, bromate ion and various organic compounds (see OZONE). The average concentration of NO_3^- found in Miami, Florida, pools is well below the drinking water standard of 45 ppm (72). The primary source of chlorate is believed to be an impurity in hypochlorite sanitizers, especially sodium hypochlorite. No drinking water standard for chlorate has been issued.

The concentration of inorganic and organic chloramines in pool water is controlled by superchlorination or shock treatment. Because chloramines are decomposed by sunlight, their effects are more noticeable in indoor pools or spas. Nitrogen trichloride, the primary volatile chloramine, is a strong irritant similar to chlorine. Its effect is noticeable at $>0.5 \text{ mg m}^{-3}$ ($>0.1 \text{ ppm}$) (73). The concentration of NCl_3 depends on the extent of ventilation and typically varies from 0.2 to 0.5 mg m^{-3} (0.04 to 0.1 ppm) (74).

THMs are formed by reaction of bromine and chlorine sanitizers not only with bather introduced contaminants, but also with ancillary chemicals. The average concentration of total trihalomethanes TTHMs ($\sim 0.1 \text{ mg m}^{-3}$, 20 ppb) in air close to the surface of pool water is two orders of magnitude below the OSHA permissible exposure limit (8-h time weighted average) for the main THM chloroform in air (10 mg m^{-3} , 2 ppm) (75). The average concentration of TTHMs in the bulk air is significantly lower. Spas, with their higher bather load, will have a higher concentration than indoor pools which will be higher than outdoor pools. The average concentration of TTHMs in pool water tend to be at or above the drinking water standard (100 ppb) (72). However, this should be tempered by the fact that swimmers or bathers do not drink pool or spa water. Saline pools have shown higher concentrations of TTHMs due to higher levels of bromoform (72).

Toxicity of Chlorine Sanitizers. Chlorine-based swimming-pool and spa and hot-tub sanitizers irritate eyes, skin, and mucous membranes and must be handled with extreme care. The toxicities are as follows: for chlorine gas, TLV = 1 ppm; acute inhalation LC_{50} = 137 ppm for 1 h (mouse) (75). The acute oral LD_{50} (rats) for the liquid and solid chlorine sanitizers are NaOCl (100° basis) 8.9 g/kg (76), 65° Ca(OCl)_2 850 mg/kg, sodium dichloroisocyanurate dihydrate 735 mg/kg, and trichloroisocyanuric acid 490 mg/kg. Cyanuric acid is essentially nontoxic based on an oral LD_{50} $> 20 \text{ g/kg}$ in rabbits. Although, it is mildly irritating to the eye, it is not a skin irritant. A review of the toxicological studies on cyanuric acid and its chlorinated derivatives is given in ref. 77.

Safe Storage and Handling of Sanitizers

Swimming-pool sanitizers should be kept in closed containers in a cool, dry area, segregated from other nonpool materials (such as ammonia, fertilizers, acids, paints, solvents, etc), and should never be mixed with each other or with other materials. Explosive NCl_3 can be generated on mixing hypochlorites and isocyanurates together (especially in the presence of water) or with nitrogen compounds (eg, ammonia, urea, quats, etc), mixing chloroisocyanurates with alkaline materials (eg, soda ash solution), or storing moist or wet chloroisocyanurates in closed or unvented containers. Mixing hypochlorites and chloroisocyanurates with pool acid (HCl) generates toxic chlorine gas. At very high temperatures, eg, in a fire, swimming-pool chemicals release toxic or hazardous gases, eg, Ca(OCl)_2 gives O_2 and Cl_2 ; sodium dichloroisocyanurate gives CO_2 , N_2 , ClCN , $(\text{CN})_2$, Cl_2 , CO , NCl_3 , and N_2O ; trichloroisocyanuric acid gives Cl_2 , N_2 , ClC(O)NCO , and trichloroisocyanuric acid vapor; and cyanuric acid gives isocyanic acid (HNCO).

BIBLIOGRAPHY

"Water, Treatment of Swimming Pools" in *ECT* 2nd ed., Vol. 22, pp. 124–134, by J. P. Faust and A. H. Gower, Olin Corp.; in *ECT* 3rd ed., Vol. 24, pp. 427–441, by J. A. Wojtowicz, J. P. Faust, and F. A. Brigano, Olin Corp.

- Swimming Pool and Spa Industry Survey, The National Spa and Pool Institute, Alexandria, Va., 1993.
- Pool-Spa Operator's Handbook*, National Swimming Pool Foundation, San Antonio, Texas, 1990, p. 11.
- Standard 50*, National Sanitation Foundation, Ann Arbor, Mich., 1996.
- The Chemical Economics Handbook*, Stanford Research Institute International, Stanford, Calif., 1995.
- R. E. Connick and Y-T. Chia, *J. Am. Chem. Soc.* **81**, 1280 (1959).
- J. C. Morris, *J. Phys. Chem.* **70**, 3798 (1966).
- J. A. Wojtowicz, *J. Sw. Pool and Spa Ind.* **1**(2), 9 (1995).
- L. O. Petterson and U. Grzeskowiak, *J. Org. Chem.* **24**, 1414 (1959).
- R. G. Rice, *J. Sw. Pool and Spa Ind.* **1**(1), 25 (1995); W. Hartwig, *J. Sw. Pool and Spa Ind.* **2**(1), 25 (1996).
- M. T. Yayha, L. K. Landeen, S. M. Kutz, and C. P. Gerba, *J. Environ. Health* **51**, 282 (1989); M. T. Yayha and co-workers, *Can. J. Microbiol.* **36**, 109 (1990).
- U.S. Pat. **4,250,910** (Feb. 1981) and U.S. Pat. **4,662,387** (May 5, 1987), L. H. King.
- U.S. Pat. **5,004,549** (Apr. 2, 1991), R. B. Wood, R. M. Mullins, and J. A. Wojtowicz (to Olin Corp.); U.S. Pat. **5,112,521** (May 12, 1992), R. M. Mullins, R. B. Wood, and J. A. Wojtowicz (to Olin Corp.).
- U.S. Pat. **4,876,003** (Oct. 1989), J. Casberg (to Olin Corp.).
- Swimming Pool and Spa Standards*, The National Spa and Pool Institute, Alexandria, Va., 1992.
- J. C. Hoff, *Inactivation of Microbial Agents by Chemical Disinfectants*, U.S. Environmental Protection Agency, EPA 600/2-86/067, 1986; NTIS PB86-232-568.
- Z. Zhang, Ph.D. thesis, University of Houston, Houston, Texas, 1988.
- M. Fair, G. C. Geyer, and D. A. Okun, *Water and Wastewater Engineering*, Vol. 2, John Wiley & Sons, Inc., New York, 1967, pp. 31–36.
- Standard Methods for the Examination of Water and Wastewater*, 19th ed., American Public Health Association, Washington, D.C., 1995.
- P. D. Lomas, *J. Assoc. Publ. Analysts* **5**, 27 (1967).
- N. A. Clarke, G. Berg, P. W. Kabler, and S. L. Chang, *International Conference on Water Pollution Research*, Pergamon Press, London, Sept. 1962.
- R. L. Valentine and C. T. Jafvert, *Environ. Sci. Technol.* **19**, 287 (1985).
- I. W. Wei and J. C. Morris in A. J. Rubin, ed., *Chemistry of Water Supply and Treatment*, Ann Arbor Publishers, Inc., Ann Arbor, Mich., 1975, p. 297; T. Palin, *Water Water Eng.* **54**, 151, 189 (1950).
- D. Nelson, *Swimming Pool Disinfection with Chlorinated-s-triazinone Products*, Special Report 6862, Monsanto Co., St. Louis, Mo., revised May 1975.
- T. Palin, *Water Water Eng.* **54**, 248 (1950); J. Fuchs, *Chem. Ztg.-Chem. App.* **86**, 76 (1962).
- W. D. Stanbro and W. D. Smith, *Environ. Sci. Technol.* **13**, 466 (1979); Z. Alouini and R. Seux, *Wat. Res.* **21**, 335 (1987).
- Z. Alouini and R. Seux, *Wat. Res.* **22**, 1519 (1988).
- W. R. Samples, Ph.D. thesis, Harvard Univ., Cambridge, Mass., 1959.
- U.S. Pat. **4,780,216** (Oct. 25, 1988), J. A. Wojtowicz (to Olin Corp.).

29. L. N. Plummer and E. Busenberg, *Geochimica et Geochimica Acta* **46**, 1011 (1982).

30. J. E. O'Brien, J. C. Morris, and J. N. Butler in A. J. Rubin, ed., *Chemistry of Water Supply and Treatment*, Ann Arbor Publishers, Inc., Ann Arbor, Mich., 1975, p. 333.

31. B. Solastiouk, thesis, Lorraine National Polytechnic Institute, Nancy, France, 1989.

32. J. A. Wojtowicz, *J. Sw. Pool and Spa Ind.* **1**(1), 7 (1995).

33. J. A. Wojtowicz, *J. Sw. Pool and Spa Ind.* **3**(1), 22 (1997)..

34. J. A. Wojtowicz, *J. Sw. Pool and Spa Ind.* **1**(3), 19 (1995).

35. R. A. Pisigan, Ph.D. thesis, Univ. of Florida, 1981.

36. F. Langelier, *J. Am. Water Works Assoc.* **38**, 169 (1946); R. F. McCauley, *J. Am. Water Works Assoc.* **52**, 1386 (1960).

37. J. A. Wojtowicz, *J. Sw. Pool and Spa Ind.*, **3**(1), 26, 35 (1997).

38. J. A. Wojtowicz, *J. Sw. Pool and Spa Ind.*, **3**(1), 33 (1997).

39. J. A. Wojtowicz, *J. Sw. Pool and Spa Ind.*, **3**(1), 16 (1997).

40. J. A. Wojtowicz, *J. Sw. Pool and Spa Ind.*, **2**(2), 23 (1996).

41. P. Brady, K. M. Sancier and G. Sirine, *J. Am. Chem. Soc.* **85**, 3101 (1963).

42. J. Gardiner, *Water Res.* **7**, 823 (1973).

43. M. Pinsky and H.-C. Hu, *Environ. Sci. Technol.* **15**, 423 (1981).

44. J. A. Wojtowicz, *J. Sw. Pool and Spa Ind.* **2**(2), 14 (1996).

45. R. Anderson, *Swimming Pool Data and Reference Annual* **32**, 86 (1965).

46. S. Stuart and L. F. Ortenzio, *Soap Chem. Spec.* **40**, 79 (1964).

47. E. Swatek, H. Raj, and R. E. Kalbus, National Swimming Pool Institute Convention, Las Vegas, Nev., Jan. 21, 1967.

48. P. Fitzgerald and M. E. DeVartanian, *Appl. Microbiol.* **15**, 504 (1967).

49. J. A. Wojtowicz, *J. Sw. Pool and Spa Ind.* **2**(1), 34 (1996).

50. F. A. Cotton and G. Wilkinson, *Advanced Inorganic Chemistry*, 5th ed., New York, John Wiley & Sons, Inc., 1988.

51. P. Adamson, M.S. thesis, Arizona State University, Tempe, Ariz., 1978.

52. P. Fitzgerald, *Appl. Microbiol.* **8**, 269 (1960); *Water and Sewage Works* **115**, 65 (1968).

53. W. Cochran and J. B. Hatlen, *Stud. Med.* **10**, 493 (1962).

54. E. Herman, *Pool and Spa News* (Aug. 21, 1996).

55. **U.S. Pat. 5,164,109** (Nov. 17, 1992), J. A. Wojtowicz (to Olin Corp.).

56. P. Fitzgerald, *Algicides*, No. PB 198130, National Technical Information Service, U.S. Department of Commerce, Springfield, Va., 1971.

57. P. Fitzgerald, *Water and Sewage Works* **121**, 85 (1974).

58. M. Palmer and T. E. Maloney, *Ohio J. Sci.* **55**, 1 (1955).

59. L. Culp and R. L. Culp, *New Concepts in Water Purification*, Van Nostrand Reinhold Co., New York, 1974, p. 84.

60. Gutcho, *Waste Treatment with Polyelectrolytes*, Noyes Data Corp., Park Ridge, NJ, 1972.

61. L. Akins, *Pool and Spa News*, (Aug. 20, 1997) p. 30.

62. F. A. Brigano and J. F. Carney, paper presented at the 84th meeting of the American Society of Microbiology, St. Louis, Mo., March 4–9, 1984.

63. *Household and Industrial Bleach Systems*, Colin A. Houston & Associates, Inc., Mamaroneck, N.Y., 1988.

64. R. S. Kebabjian, *J. Environ. Health* **58**, 8 (1995).

65. C. P. Gerba and P. Gerba, *J. Sw. Pool and Spa Ind.* **1**(3), 9 (1995).

66. *Suggested Health and Safety Guidelines for Public Spas and Hot Tubs*, U.S. Department of Health and Human Services Publication No. 99-960, Centers for Disease Control, Atlanta, Ga., 1985.

67. *Morbidity and Mortality Weekly Report* **28**, 182 (1979).

68. J. Washburn and co-workers, *J. Am. Med. Assoc.* **235**, 2205 (1976).

69. F. Sausher and co-workers, *J. Am. Med. Assoc.* **239**, 2362 (1978).

70. *Morbidity and Mortality Weekly Reports* **30**, 329 (1981).

71. E. W. Mood, C. C. Clarke, and A. Gelperin, *Am. J. Hyg.* **54**, 144 (1951).

72. J. A. Beech and co-workers, *Am. J. Public Health* **70**, 79 (1980).

73. M. Hery and co-workers, *Cahiers de notes documentaires No. 156* (1994).

74. H. J. Jessen, *Z. Gesamte Hyg.* **32**, 180 (1986).

75. *Registry of Toxic Effects of Chemical Substances*, Vol. 2, U.S. Dept. of Health and Human Services, Washington, D.C., 1980 ed.

76. Technical data, Industrial Biotest Labs, Northbrook, Ill., 1970.

77. B. G. Hammond and co-workers, *Env. Health Perspect.* **69**, 287 (1986).

General References

C. White, *Handbook of Chlorination*, Van Nostrand Reinhold Co., New York, 1986.

W. Stumm and J. J. Morgan, *Aquatic Chemistry*, 3rd ed., John Wiley & Sons, Inc., New York.

John A. Wojtowicz
Consultant

WATERPROOFING AND WATER/OIL REPELLENCY

Principles of Repellency

A thorough discussion of water repellency was published in 1963 (1).
When a drop of liquid is placed on a solid surface, the shape of the drop depends on the equilibrium among three forces, as shown in Figure 1.

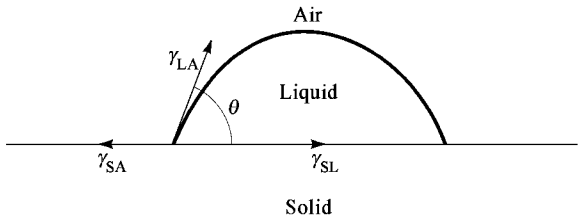


Fig. 1. Diagram of forces determining shape of a drop of liquid placed on a solid surface. See text.

On an ideal surface, the contact angle θ depends on three surface tensions: the liquid–air interfacial tension, γ_{LA} ; the solid–air interfacial tension, γ_{SA} ; and the solid–liquid interfacial tension, γ_{SL} . Wetting of the solid by the liquid results from the reduction of the contact angle of the advancing liquid so that the liquid spreads easily. A surface is made repellent by raising the advancing contact angle, so that spreading and migration into capillaries do not occur. The nonequilibrium spreading coefficient S_{SL} for a liquid on a solid is defined by the equation (2):

$$S_{SL} = \gamma_{SA} - (\gamma_{LA} + \gamma_{SL})$$

The term γ_{SA} is the same as the free energy of the solid surface in air, and γ_{SL} is the free energy per unit area of the solid–liquid interface. If the spreading coefficient is positive, spreading can occur spontaneously and the liquid should spread to form a thin film on the solid. A liquid spreads on a surface because spreading causes a decrease in free surface energy. Conversely, spreading does not occur if the spreading coefficient is negative. This is because γ_{SA} is low, ie, the surface is a low energy solid or has been treated with a chemical that converts it to a low energy solid. Surfaces that contain nonpolar hydrocarbon, siloxane, or fluorocarbon groups are low energy surfaces. These repel a liquid unless the surface tension of the liquid, γ_{LA} , and the interfacial tension of the solid–liquid interface, γ_{SL} , are sufficiently low to make the spreading coefficient positive.

The measurement of contact angles is a convenient means for judging the degree of wetting or repellency for a liquid on a planar solid surface. Measurements of advancing or receding contact angles, which form when the liquid moves outward or contracts on the surface, are generally more reproducible than measurements of static contact angles (2) because of the existence of a large number of metastable states at the three-phase boundary (3). Methods for measuring advancing and receding angles are available (2). The advancing contact angle most nearly represents the low surface energy or repellent component of the surface, whereas the receding angle is representative of the high surface energy or more wettable component of the surface. Most real surfaces are made up of a complex mixture of low and high energy surface components and exhibit significant differences between advancing and receding contact angles. This phenomenon is termed contact angle hysteresis. Whereas receding angles can be important, the advancing contact angle is appropriate for investigating whether a liquid will advance or "wet" over a substrate, and the bulk of this discussion will focus on advancing contact angle.

A number of methods have been developed for estimating advancing contact angles of liquids on surfaces. These methods can be used to estimate wettability of a given liquid–solid system.

The critical surface tension (CST) of a surface is the maximum surface tension for a liquid that has an advancing contact angle θ_a equal to 0° . Advancing contact angles on low energy solids depend approximately on only the surface tensions of the wetting liquids (4). The CST for a planar surface can be estimated by measuring the advancing contact angles of a series of liquids, eg, alkanes, on the surface. A plot of cosines of advancing contact angles against surface tensions of the liquids gives approximately a straight line, as shown in Figure 2.

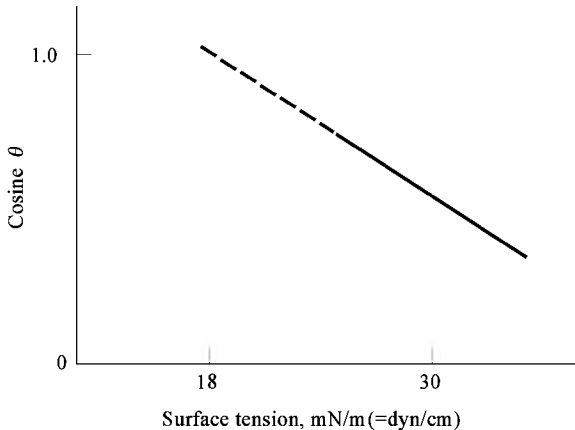


Fig. 2. Determination of a critical surface tension.

Extrapolating this line to $\cos\theta_a = 1$ (contact angle $= 0^\circ$) gives the surface tension required for a liquid to wet the solid. Given a surface of known CST and a liquid of known surface tension, an estimate of advancing contact angles can be made.
There are several other methods for estimating the contact angles that specific liquids will make on surfaces. These methods depend upon breaking

the free surface energy into different components (5). A surface can be characterized as having a polar and nonpolar surface energy component which together equal the total surface energy. A specific liquid also has a polar component and a dispersive (nonpolar) component of the surface tension, which interact with the surface. Several theories allow the advancing contact angle to be predicted based upon this approach. In many cases the advancing contact angles are quite accurate and can be predicted, given the polar and dispersive parameters (5–9). Another approach is to use the acid–base interaction to predict advancing contact angles based upon the inherent acid and base character determined for a surface and its contacting liquids (10).

A repellent material functions by lowering the surface energy to a value such that the liquid now has a high advancing contact angle on the surface. For example, a fluorochemical finish on a nylon fabric can lower the surface energy from ca 43 mN/m (dyne/cm) for the untreated fabric to <10 mN/m for the treated fabric. Such a treated fabric would repel water with a surface tension of 72 mN/m and oils or solvents with surface tensions down to ca 20 mN/m. Unfortunately, advancing contact angles cannot be measured directly on textile yarns that are complex bundles of twisted fibers. Other methods must be employed to predict the wettability behavior.

Most of the surfaces that require repellent treatments are not smooth but contain capillaries into which a liquid can migrate, even though the advancing contact angle of the liquid on the surface is >0°. The law for the movement of liquids into an idealized capillary is given by the equation:

$$\Delta P = \frac{2\gamma_{LA}\cos\theta_a}{r} - \frac{2(\gamma_{SA} - \gamma_{SL})}{r}$$

where ΔP is the pressure required to force the liquid back out of the capillary, and r is the diameter of the capillary. ΔP is positive, ie, the liquid rises in the capillary, if the liquid wets the capillary with an advancing contact angle $\theta_a < 90^\circ$ ($\cos\theta_a > 0$). If $\theta_a > 90^\circ$, $\cos\theta_a$ is negative, making ΔP negative. In the latter case, the solid repels the liquid, and pressure is necessary to force the liquid into the capillary. The second equation for ΔP shows that, for a negative value of ΔP , γ_{SA} (the free energy or surface tension of the solid in air) must be less than γ_{SL} (the interfacial tension of the solid in the liquid).

The limiting contact angle for repellency of a solid surface for a liquid depends on the physical form of the solid. The capillaries in textiles, paper, leather, or masonry are not ideal, smooth, and round. In a textile or in paper, the wicking of liquids can be lateral, ie, through the sides of fibers, or longitudinal, ie, along the length of fibers (11–13). In masonry, liquids can migrate through an array of irregular capillaries.

In a bundle of fibers, as in a textile yarn, fabric, or paper, the fibers may repel a liquid even though the advancing contact angle is <90° (9). Pressure on the liquid that contacts the fibers, however, can force the liquid into capillaries or interstices between fibers, thereby effecting wicking. The pressure required for wicking into repellent treated fibers depends upon the surface energy of the treated fibers and the geometry of the system. Wicking occurs even without pressure on the liquid if the advancing contact angle is approximately 70°; wicking occurs because the liquid surface can contact fibers beneath the surface. The combination of lateral and longitudinal wicking into a fiber assembly illustrates the gradual wicking of aqueous fluids, solvents, or oils into fabrics, paper, or nonwoven products (see NONWOVEN FABRICS). The gradual advance of a moving liquid front along capillaries can also explain the migration of liquids through masonry and concrete.

Textiles

Early waterproofing of textiles was generally accomplished by the use of impermeable coatings, such as natural fats, oils, waxes, pitch, and asphalt. Later, vulcanized natural rubber became an important waterproofing material. By the 1930s, repellent finishes durable to washing and dry cleaning were introduced. The following definitions apply to fabrics with water-repellent properties (14).

Water-repellent fabrics resist wetting or repel waterborne stains; they pass AATCC Test Method 22 (Spray Test).

Water-resistant fabrics protect against water penetration during a light or brief shower and pass AATCC Test Methods 22 and 42 (Impact Penetration Test).

Rain-resistant fabrics protect against water penetration during a rain of moderate intensity and pass AATCC Test Methods 22 and 35 (Rain Test).

Waterproof is normally used to describe plastic, plastic-coated, or nonbreathable fabrics. However, the term has been used for some chemically coated fabrics that allow the penetration of air.

Durable finishes maintain a high level of performance after laundering, dry cleaning, or both. Durable has also been used to describe resistance to abrasion.

The levels of water repellency required to pass the preceding tests depend upon the specifications for particular fabrics.

Oil-repellent fabrics resist wetting by oily liquids and repel oilborne stains. The level of performance of such fabrics is judged by AATCC Test Method 118.

Waterproof Finishes. Waterproofing results from coating a fabric and filling the pores with film-forming material such as varnish, rubber, nitrocellulose, wax, tar, or plastic. The materials may be applied as hot melts, eg, waxes or some polymers, as solvent solutions, or as aqueous latexes. The continuity of the film provides the water resistance. Except for tents, tarpaulins, and covers, coated fabrics have been largely replaced by plastics, and by fabrics treated with water and oil repellents that do not reduce permeability to air and water vapor. Fabrics are also commonly laminated to films, such that the total structure is waterproof (15), or in some cases water-resistant but breathable (16).

Repellent Finishes. The following are six classes of repellent textile finishes.

Fluorochemicals. Fluorochemicals are the most important class of repellents for textiles. They are the only repellents that provide repellency to water, waterborne stains, oil, oilborne stains, and oily particulates. The various products have a variety of repellency and durability properties for certain fabrics, and the specific compositions are proprietary. The first company to market fluorochemical repellents was 3M in the 1950s (Scotchgard Fabric Protector), followed by DuPont (Zepel and Teflon Fabric Protectors). Several other companies such as Auralux Corporation, Ciba Specialties, Eastern Color & Chemical, Glo-Tex Chemicals, IVAX Industries, Lindley Laboratories, NICCA U.S.A., Piedmont Chemical, Sedgefield Specialties, Sequa Chemicals, and Yorkshire Pat-Chem also market such finishes (17). Fluorochemical finish application areas include rainwear, upholstery, drapery, and automotive fabrics, roofing materials, and carpeting. Both natural and synthetic textile fibers can be treated.

The performance of fluorochemical repellents depends on the presence of perfluorocarbon chains, $CF_3(CF_2)_N$. The length of the perfluoroalkyl group is at least four carbon atoms. For many products, it is eight carbon atoms, or a mixture of chain lengths averaging about eight. The products are primarily copolymers of perfluoroalkyl acrylates or methacrylates with other comonomers. The comonomers in the polymeric products are esters of acrylic or methacrylic acid containing alkyl groups of several carbon atoms, alkylamide groups, or polyether groups. The comonomers modify the physical properties of the polymers (eg, control the hand of the finished fabric) or improve the performance of the polymer (eg, provide soil release) and reduce the manufacturing cost of the product.

Fluorochemicals repel both water and oil because they produce an extremely low energy surface (18–26). The effectiveness of the fluorochemicals depends upon uniform surface coverage and orientation of the molecules on the fiber surface so that the perfluoroalkyl chains are directed away from the surface. The result is a CST as low as 5–10 mN/m (dyne/cm). Fluorochemical finishes are often formulated with nonfluorinated resin-based water-repellent extenders. These water repellents not only reduce the cost of the finish but may also improve durability (27,28).

Quarapel is an important combination of fluorochemical finish and resin-based extender developed by the U.S. Army Natick Laboratories for military use. This finish typically contains 4–6 wt % commercial fluorochemical emulsion, 4–6 wt % resin-based repellent emulsion, 0.1 wt % acetic acid, and 5 wt % isopropyl alcohol. If necessary, the formulation includes a catalyst to cross-link the resin-based component. Quarapel specifications demand excellent initial water and oil repellency and excellent durability to washing and dry cleaning.

Fluorochemical repellents are commercially available as emulsions or solvent solutions. The most widely used emulsions for fabrics and carpet are cationic, but nonionic emulsions are becoming more prevalent. The emulsifier in the formulation can affect the repellency and durability of the product (28). Surfactants used in the fluorochemical emulsions or added to finish baths should be nonrewetting and have a minimum adverse effect on oil repellency. Solvent solutions of fluorochemicals are becoming less common as a result of toxicity and environmental concerns.

The choice of application conditions for fluorochemical repellents depends on the textile, the repellent formulation, and level of repellency required. Suppliers provide detailed recommendations in their product bulletins. Where a high level of repellency is required or the generally somewhat higher cost is

tolerable, pad application is the most effective method. For lower levels of repellency, spraying or foam finishing may suffice. The usual application level of repellent solids on any textile is 0.2–0.5% of the fiber weight, although higher levels may be needed. There should be no silicones on fabrics prior to treatment with fluorochemicals, because trace levels of silicone lubricants can diminish the oil repellency of the fluorochemical. The effect may result from poor wettability of the silicone-treated fiber with the fluorochemical repellent. Also, silicones can attract oils and oily soil.

Many fluorochemical finishes for fabrics require curing at up to about 175°C. Curing allows melt-spreading of the fluorochemical to ensure maximum leveling of the finish on the fibers and to promote optimum orientation of the pendant fluorinated portion of the molecules on the fibers.

Nonwoven textile products often require fluorochemical finishes (29). Medical nonwovens (operating room gowns, central supply wraps) represent the largest market. Although they require oil and alcohol holdout, aqueous-based fluid repellency is the most important property. Because these products are mostly disposable, the treatments are designed for nondurable, high water repellency. This is best achieved on nylon, polyester cellulose and adhesively bonded cellulose nonwovens with a low level of fluorochemical, and a high level of a combination wax resin emulsion or water-repellent extender. The polypropylene nonwovens, fluorochemical treatments must cure at low temperature conditions, and sometimes require special wetting agents.

Silicones. Silicones are exceeded only by fluorochemicals in the volume used as repellents for textiles. They are widely used on cellulosic and synthetic fiber fabrics. Silicones provide water-based stain resistance; good durability to washing; improved tear strength; a soft, slick hand; and improved fabric sewability.

The most widely used silicones are polymers of methyl(hydrogen)siloxane and of dimethylsiloxane. Polydimethylsiloxane is the basic polymer used in silicone repellents. If the polymer is terminated with methyl groups it is inert; however, if it is terminated with hydroxyl groups, it can be cross-linked. Continuous, durable coatings result from the use of curable blends of polydimethylsiloxane and polymethyl(hydrogen)siloxane. The silicone finish encapsulates individual fibers.

Silicones are supplied as aqueous emulsions or as solvent solutions. Dow-Corning and OSi Specialties are primary manufacturers and suppliers. Emulsions are usually applied to fabrics by padding or exhaustion. Solvent solutions can be applied by spraying. With either type of product, coapplication of a catalyst is necessary. The level of silicone solids on the weight of fabric should be 0.5–1.5%. Most of the silicone emulsions can be coapplied with durable-press resins. Curing occurs at about 150°C.

The water repellency of silicone finishes results from the low CST (ca 22 mN m or dyne cm) produced by the methyl groups in the silicone that are oriented away from the fiber surface. The CST is lower than that produced by any class of compounds except for fluorochemicals.

Resin-Based Finishes. Resin-based water repellents are durable finishes that are modified melamine resins, blended with waxes. In some cases, the resin helps disperse the wax in the repellent formulation. The resin provides water repellency and binds the wax onto the fabric.

Resin-based repellents may be used alone or in combination with durable-press resins. They are widely used as extenders for fluorochemical repellents. When used alone, several of the resin-based finishes require an acid catalyst and curing at temperatures above 150°C for maximum repellency and durability. When coapplied with durable-press finishes, which themselves require a magnesium chloride catalyst, the catalyst and curing conditions for the durable-press finish provide the necessary conditions for the repellent.

Resin-based finishes are applied to fabrics by padding. A concentration of 1.5% resin solids on the weight of the fabric is a common application level. The concentration of repellent solids required for good initial repellency and good durability is about 0.3–0.5%, based on the weight of the fabric. The compositions of commercial products are proprietary.

Waxes and Wax–Metal Emulsions. Waxes and wax–metal emulsions are the lowest priced, widely used water repellents and fluorochemical extenders. They can be applied by padding or exhaustion with no cure commonly required. However, waxes have only poor-to-fair durability to washing and dry cleaning, and tend to show streaks from abrasion.

Several types of wax and wax–metal emulsions are water repellents (30,31). Among these are wax dispersions without metal salts and wax dispersions containing aluminum or zirconium salts. The products that do not contain metal salts are anionic emulsions of wax, used alone or in combination with durable-press resins. Specific compositions are proprietary. Their chief use is on nylon, polyester, and acetate fabrics.

Wax–metal emulsions are useful on a variety of fabrics, alone or in combination with resins. Wax–aluminum emulsions, which have largely lost their market share to wax–zirconium emulsions, usually contain paraffin wax, aluminum acetate or formate, and a dispersing agent and protective colloid, eg, glue. They tend to develop amine odors when used with thermosetting resins. Wax–zirconium salt emulsions are useful on a wide variety of natural and synthetic fibers. Some finishes have considerable durability to laundering.

Organometallic Complexes. Werner-type complexes of chromium and long-chain carboxylic acids, eg, stearic acid, are water repellents for fabrics of natural and synthetic fibers. The complexes have a small market in the textile industry.

Modification of Textile Fibers. The reaction of hydrophobic chemicals with textile fibers offers the possibility of permanent repellency without alteration of the other physical properties of fibers. However, the disadvantages caused by complex processing, and resultant higher costs of carrying out chemical reactions on fiber in commercial textile plant operations, have limited the commercial applications. The etherification and esterification of cellulose have been most effective in terms of achieving durable water repellency (32,33). Radiation grafting of reactive repellents onto fibers has been studied as a potential commercial process (34,35), as has modification by plasma polymerization of gas monomers or plasma initiated polymerization of liquid monomers (36).

Fabric Construction for Water Repellency. Fabric construction, including twist, ply, and coarseness of yarns, affects the performance of water repellents. Waterproof films can more easily be formed on close weaves than on open-weave fabrics. Hydrophobic finishes, which make individual fibers repellent without altering fabric porosity, are generally applied to fabrics whose pores are small (37). The relation of rainwear fabric construction to the performance of repellents has been reviewed (38). Some reports indicate that fabric roughness reduces repellency (28,37). Mechanical action on fabrics, even after treatment, can reduce repellency if the action increases fiber roughness or exposes fibers that have little repellent treatment.

Carpet. Carpet, an important textile, may also be treated to provide water and oil repellency; however, the principal functions of the current carpet treatments are to provide soil and stain resistance. High quality carpets, especially those made from nylon, polyester, or wool, have a significant proportion of the surface coated with fluorochemical materials. The treatments can be spray-applied to a finished carpet or applied directly to the fiber during the spinning or dyeing operations. Suitable fluorinated resin materials are readily available from 3M or DuPont.

Health and Safety Factors. The Material Safety Data Sheets provided by the suppliers should be consulted for each product. In general, products are aqueous emulsions with low levels of toxicity. Products with high solvent content have mostly been eliminated. Personnel handling the chemicals should always avoid contact of the products with skin and eyes, and avoid exposure to vapors if the product contains volatile components.

Test Methods. Many tests are useful for determining water and oil repellency in textiles and for testing the durability of finishes to washing and dry cleaning. The most widely used tests in the United States are described in the *Technical Manual* of the American Association of Textile Chemists and Colorists (14). The selection of tests depends upon the textile being tested, ie, whether it is a fiber, fabric, or nonwoven product. The selection also depends on the nature of the finish, ie, whether it is waterproof, water-repellent, water- and oil-repellent, or solvent-repellent. Special tests may be necessary for fabrics or garments that must meet specific requirements, such as repellency of shell and lining for lined garments (39). Several supplementary tests are important, depending on the textile. These include air and water-vapor permeability, repellency to specific chemicals or vapors, fabric stiffness or softness, color, effects on dyes, and durability of physical properties and color to exposure to light or atmospheric gases.

In the following descriptions of some widely used repellency tests, the term fabric designates any woven, knitted, or nonwoven textile.

The spray test is one of the most commonly used tests for fabrics and nonwoven products. The test material is held tightly on an embroidery hoop and mounted at a 45° angle to the horizontal and 15 cm below a spray nozzle. The fabric is sprayed with 250 mL of water. The degree of repellency is rated by comparing the sprayed fabric or nonwoven with pictures on a standard chart (AATCC Test Method 22 (7); INDA Standard Test 80.1-92 (40)). Durability of the finish to dry cleaning can be evaluated by first cleaning the fabric according to AATCC Method 86 (14).

The rain test simulates the effects of rainfall; the hydrostatic head on the spray controls the intensity of spraying. The repellency is rated by the weight of water that penetrates the fabric and is absorbed by a blotter mounted behind the fabric at a specific intensity of spraying (AATCC Test Methods 35 and 42; INDA Standard Test 80.2-92).

Oil repellency is measured by observing a fabric's resistance to wetting by a selected series of numbered test liquid hydrocarbons with a range of surface tensions. The fabric rating is based on the liquid that does not wet the fabric surface in a specified time (AATCC Test Method 118 and INDA

Standard Test 80.7-92). Suppliers of fluorochemical repellents may recommend different times or a different series of hydrocarbons in the evaluation of treated fabrics.

Alcohol holdout tests, which are also used to measure aqueous fluid repellency, involve placing drops of aqueous isopropyl alcohol solutions of concentrations 10, 20, ... 100 wt % on a fabric surface. The rating for the fabric is based on the most concentrated solution that does not penetrate the fabric in the specified time frame (3M Water Repellency Test II, Water Alcohol Drop Test (41); INDA Standard Test 80.6).

The nonwovens industry also uses a saline repellency test, especially for medical fabrics. Fabrics are evaluated for the lapsed time before wetting when a 115-mm column of water in a mason jar is applied to the surface (INDA Standard Test 80.5-92).

The hydrostatic-pressure test is performed on fabric mounted under the orifice of a conical well. The fabric is subjected to increasing water pressure at a constant rate until leakage occurs at three points on the fabric's undersurface. The rating is the height of the water head in centimeters above the fabric (AATCC Test Method 127; INDA Standard Test 80.4-92).

The dynamic absorption test measures the resistance of fabrics to wetting by water, not the repellency of the total fabric surface. A weighed portion of fiber, yarn, or fabric is tumbled in water for 20 minutes; it is then removed and reweighed to determine the percentage of water absorbed (AATCC Test Method 70).

The specifications for water, oil, solvent, or soil repellency vary with the use of the textile, the nature of the textile, and the nature of the repellent (27–29,42,43). Suppliers of repellents may specify a minimum level of repellency for the approved use of a repellent.

Leather

For fashionable leather apparel, high quality upholstery such as nubuc or aniline, and for shoes, oil- and water-repellent treatments preserve the leather's original appearance and provide for easy care. The main characteristics required by easy-care leather are the ability to shrug off water- and oil-based stains, and to resist soiling. Water-resistant treatments are useful to reduce the tendency of leather to become stiff and uncomfortable after wetting and drying. Various approaches have been made over the years to meet this need. All have their advantages and disadvantages and none can be said to offer the ultimate solution.

Before the introduction of synthetic water repellents, the best way to protect leather was to apply large concentrations (up to 50% of the leather weight) of oil, tallow, or grease. The treated leather was uncomfortable because the pores of the leather were filled and the water-vapor permeability was low. Water repellents available today are useful at low concentrations and make the leather surface repellent without reducing permeability to water vapor. The theoretical and practical aspects of water repellents for leather have been reviewed (44). Water should bead on the surface of water-repellent leather because of treatment of the individual fibers, both on the surface and in the interior of the leather. The fibers should not become wet even with prolonged soaking, and the leather should be permeable to water vapor for comfort.

Commonly used repellents for leather are silicones, chrome complexes of long chain fatty acids, and fluorochemicals. Fluorochemical repellents also provide repellency to oils and greases so that the treated leather resists staining. A water repellent may also be a hydrophobic chemical insolubilized in the leather. A simple water-repellent treatment consists of forming an aluminum soap in leather by the two-step process of applying a soap, and then an aluminum salt.

The silicones, as supplied by Dow Corning and OSi Specialties, are an important class of repellents (44,45). Numerous patents describe the use of silicones for leather (46,47). Silicones are linear polymers, eg, dimethylsiloxane polymers, that bond to the leather. They can be applied as aqueous dispersions in the presence of glutaraldehyde acting as a bridge between the silicone and amide groups of the leather (48). Silicones can also be applied to leather as solvent solutions in a post-tanning operation. The concentration of silicone used is 4.5–15 wt % to provide repellency that resists extended flexing (45,48).

Organic titanates can be used to cross-link silicones on leather (49). Tetrabutyltitanate and tetrakis(β-aminoethoxy)titanium cross-link silicones, eg, poly(dimethyl siloxane), to increase both water repellency and durability. For example, good repellency results from the impregnation of leather with a solution of 10 wt % tetrabutyltitanate in butyl acetate, followed by impregnation with a 9:1 mixture of silicones and tetrabutyltitanate (50).

Chrome complexes of stearic and myristic acids provide water and aqueous stain resistance, dimensional stability, and lubricity. The products may also enhance the appearance and durability of leather. The chrome complex reacts with the leather molecules to form a permanent bond.

Chrome-complexed fluorochemicals, as well as fluoropolymers, are widely used products. The compositions are proprietary. Fluorochemicals provide a high degree of water repellency as well as repellency to aqueous stains, oils, grease, and oilborne stains. Traditionally, treatments are applied during a drum process in which about 30 min are required for full penetration of the leather to occur. Products are also available for application with spraying equipment and roll coaters.

Test Methods. Tests for measuring the water and oil repellency of leather include ASTM and AATCC tests, as well as tests developed by suppliers.

- In the Spray Test (AATCC Test Method 22), water is sprayed on a taut surface and rating is based on comparison with a standard chart (14).
- In Oil Repellency (AATCC Test Method 118), drops of oils of various surface tensions are placed on the leather and monitored to absorption (14).
- In the 3M Water Alcohol Drop Test (41), this test is designed to provide a simple, rapid method to assess the aqueous stain resistance of substrates treated with a protective finish.

The 3M Abrasion Test (51) is used to assess the durability of a protective fluorochemical finish by evaluating its resistance to abrasion and wear. The surface is abraded with an AATCC crockmeter fitted with sandpaper.

For dynamic water resistance, the Bally Penetrometer Test, IUP 10 (52) measures water penetration.

For dynamic water resistance, the Maeser Test, ASTM D2099 (53) measures the number of flexes and water penetration. The flex imparted to the leather is a magnification of the flex given the vamp of a shoe in actual wear.

For static water absorption (ASTM D1815) (53), the weight of water absorbed during 30 min, 2 h, or 24 h, with all surfaces of the leather exposed to the water, is measured.

Paper and Paperboard

Paper or paperboard that repels water may be designated waterproof or water-resistant. In this context, waterproof means resistant to the penetration of both water and water vapor. A chemical additive, called a sizing agent, can provide water resistance by lowering the surface energy of the paper without greatly affecting the porosity. A water-vapor barrier or moisture-vapor barrier (MVB) is provided by a film or film-forming coating used in a manner to seal the surface. Acrylic-based overprint varnishes are often used to provide water resistance, as well as abrasion resistance, to a printed surface. Paper or paperboard can also be made resistant to oil-based fluids. Food oil resistance is the most common application for this property. To lower the surface energy enough to provide oil repellency, a fluorochemical additive is required (for pigment or clay-coated paper or paperboard, see PAPER; PAPERMAKING ADDITIVES).

Waterproof. Waterproof paper or paperboard is commonly made by lamination of a polyethylene film or wax that is applied by extrusion coating. The MVB properties derived from wax emulsions are generally inferior to those derived from extrusion coatings, but both forms provide excellent waterproof properties. Polyester film is used in food board and other applications where the composite material must withstand temperatures such as those encountered when re-heating or cooking foods in an oven. Currently (ca 1997), there are efforts to replace the standard film laminants with aqueous coatings. An important consideration is improvement in the yield and ease of repulping and recycling. Aqueous wax-based coatings are not very useful in this regard because wax residues in the recycling process cause problems in papermaking. Polyvinylidene chloride-based aqueous coatings provide excellent MVB properties, but paper and paperboard with chlorine-containing coatings are becoming less common because of concerns regarding cost, aging characteristics, and residual chlorine-containing products in the environment. Aqueous coatings that provide a good moisture barrier can be applied by air knife, rod, or rotogravure. To provide protection to exposed cut edges or to protect from failure at pinholes, it may be beneficial to use paper or board that is sized or has an oil repellent treatment. As an example, paperboard for milk carton stock is strongly sized with alkyl ketene dimer (AKD) and laminated

with a polyethylene to make it waterproof. In these ways, useful paper products can be made for boxes, bags, liners, roofing material, and protective coverings that are waterproof.

Water-Resistant. Paper or paperboard that is resistant to water can be produced by chemical addition of a sizing agent at several points in the manufacturing process or during subsequent converting operations. If the sizing is added to the aqueous suspension of pulp fibers before sheet formation, it is called internal or beater sizing. This gives chemical modification of the fiber surface throughout the sheet or ply. Surface sizing done on a paper machine can actually be a deep treatment from a pond size press, or a shallow treatment applied by a premetered size press, blade, or rod coater. A surface treatment can also be applied at a printing station in a converting operation. Over 75° of all paper and paperboard produced in North America is sized to some extent, the two largest segments being fine paper and unbleached kraft. Extensive reviews of sizing are available (54–58) as well as listings of available commercial products (59,60).

Internal Sizing. The most widely used internal sizes are alkyl ketene dimers (AKD), alkenylsuccinic anhydrides (ASA), and rosin-based sizes that are used with papermaker's alum (aluminum sulfate with 14 waters of hydration), polyaluminum chloride (PAC), or polyaluminum silicosulfate (PAS) (61). The rosin-based sizes are used under acidic conditions. Since the mid 1980's there has been a steady conversion from acid to alkaline paper production, resulting in static to declining demand for the rosin-based sizing systems. Rosin is a complex mixture of compounds and consists primarily of monocarboxylic acids with alkylated hydrophenanthrene structures (62). A main constituent of wood rosin, gum rosin and tall-oil rosin is abietic acid. Rosin is commonly modified with maleic or fumaric acid to improve efficiency. Since the 1970s, dispersions of unsaponified rosin have become more popular as a result of their improved sizing efficiency, lower alum requirements, and reduced pH sensitivity vs saponified rosins. Cationic dispersed rosin size, which can be effective at near-neutral and neutral papermaking conditions, is also available (63–65). Commercially available rosin sizes include Pexol, Neuphor, and Hi-pHase (Hercules Inc.), Plasmine and NeuRos (Plasmine), Stafor (Westvaco), Novaplus, and Novasize (Georgia Pacific), and NeuRos and Roscol (Akzo Nobel).

The AKD and ASA sizes are used in neutral to alkaline papermaking. ASA has the advantages of faster development of water-repellent properties and a lesser tendency to give slip, but it is supplied as a two-component system that must be emulsified shortly before use at the paper mill site, owing to the short active half-life of the ASA. Commercially available ASA sizes include Accosize (Cytec), Nalsize (Nalco), and Bersize (Bercen). AKD is supplied as a one-part system as an aqueous emulsion. It has an active half-life measured in months at room temperature, but this can be extended with refrigerated storage. AKD-sized paper exhibits excellent resistance to penetration by acidic fluids such as milk and fruit juice, as well as water. Commercially available AKD sizes include Hercon and Aquapel (Hercules Inc.), Rasiofob (Raisio), Keydime (Akzo Nobel), Basoplast (BASF), and Darasize (Grace Dearborn). In 1995, a new alkenyl reactive size, under the name Precis (Hercules Inc.), was introduced. It is best suited for fine paper grades where properties such as improved printability are desired since it does not give the strong sizing of the AKD and ASA.

Surface Sizing. Surface sizing is generally used for modification of other properties of paper or paperboard such as printability, smoothness, porosity, coefficient of friction, opacity, surface strength, anti-linting or coating holdout. Anionic starch is perhaps the most common additive or co-additive used for surface sizing.

Most of the surface sizes used in North America are modified styrene maleic anhydride (SMA) copolymers. Commercially available materials include Scripset (Monsanto Hercules Inc.), Cypres (Cytec), Sursize (Akzo Nobel), MSA (Morton), NovaCote (Georgia Pacific), and HTI (Hopton Technologies). Styrene acrylate emulsions that are commonly used include Jetsize and Unibond (Akzo Nobel), Basoplast (BASF), and Cypres (Cytec). Other materials used as surface sizes include acrylonitrile acrylate copolymer (Basoplast, BASF), stearylated melamine resin (Sequapel, Sequa), polyurethane (Graphsize, Vining Chemicals), and diisobutylene maleic anhydride copolymers (Baysynthol, Bayer).

Chromium complexes of long-chain fatty acids are excellent water repellents which are also used for their food-release properties in certain packaging applications. The presence of chromium has raised environmental concerns, despite the fact that the metal is in the trivalent rather than in the highly toxic hexavalent state. This material is available as Qulon (DuPont).

In most cases, some internal size is still required to control the depth or penetration of the surface size, the runnability through a size press, or other properties. Surface sizing reviews and leading references are available (66,67).

Oil Repellent. Fluorochemicals are the only class of material that can provide oil repellency without altering the porosity of the paper or paperboard. Physical barriers to oil penetration are used primarily for their moisture- or gas-barrier properties, with retarded oil penetration as a secondary benefit. The most common oil-repellent additives are long-chain perfluoroalkyl phosphate salts of ammonia or diethanol amine. Commercial sources include Scotchban (3M), Zonyl (DuPont), and Lodyne (Ciba Specialties). There are also a fluorochemical carboxylate salt, Lodyne (Ciba Specialties), and fluorochemical copolymers, eg, Scotchban (3M). The widest range of oily fluid holdout is provided by the fluorochemical copolymers.

Oily-based fluids become more aggressive in staining packaging papers as they become higher in polar components such as fatty acids, surfactants, and water. Fluorochemicals are primarily used in surface applications made at the size press of a paper machine or the calender stack of a board machine. To a lesser extent they are applied off-machine, in a converting operation. They have also been used in pigment, clay, or MVB coatings. Internal applications (beater addition) of the fluorochemical phosphate salts require the use of cationic retention aids to attach these anionic species to the fibers. The packaging of many items such as snack food, french fries, convenience food, and pet food is facilitated by the oily-stain protection provided by fluorochemicals.

FDA Regulations. Most of the sizing agents, waterproof coatings, and oil repellents mentioned are regulated by the U.S. Food and Drug Administration (FDA) for use in paper or paperboard that comes in direct contact with food. To avoid raising contamination issues, paper and paperboard mills often require all chemicals to be FDA-regulated if any of their products are used in food contact applications. The regulations often include limitations on the intended use, method, or point of application; upper limits on the amount that can be present on the paper or paperboard; and guidelines related to the filling, storage, and final use of the paper or paperboard (68). For that reason, it is important that every link in the production chain, from the supplier of the chemical additive to the paper mill to the package producer, know the specific limitations governing the materials used to modify the paper or paperboard when they are intended to come in direct contact with food. Suppliers of repellent additives for paper can provide information as to the FDA compliance of their materials and specific limitations surrounding their use.

Test Methods. The test methods used for water or oil repellency are quite varied, in accordance with paper or paperboard type and end use. Two tests commonly used to measure water repellency, the Hercules Size Test and the COBB Test, as well as others have been summarized (55,69,70). The Kit Test (T-559 pm-96) can be used to measure oil repellency when there is no physical barrier to oil penetration such as that provided by a film, foil, or waterproof coating (71). Another measure of oil repellency is the Turpentine Test (T-454 om-89) (71). If a physical barrier is present, tests that measure the weight of oil pick-up or show-through under conditions that simulate the end use are appropriate.

Concrete and Masonry

Water resistance is an important factor in concrete and masonry construction for the safety, health, and comfort of building occupants (see CEMENT). Several texts on concrete construction describe the methods for obtaining water resistance (72–76). The term waterproof describes concrete and masonry that is completely impervious to water and its vapor, whether or not the water is under pressure. Waterproof construction involves the use of some type of barrier that covers all surface pores or capillaries. Water repellent describes concrete or masonry that repels water without significantly reduced permeability to water vapor. In this discussion, concrete and masonry are used synonymously.

Several problems related to excess moisture in concrete are expansion, shrinkage, and cracking; efflorescence, staining, and mildew; spalling due to freeze thaw cycling; increased thermal and electrical conductivity; chemical attack; damage to contents from leakage; damage to finished walls; and corrosion of reinforcing steel. Care in choosing aggregate and in selecting an appropriate water-cement ratio leads to a watertight concrete that may not require additional waterproofing or water-repellency treatment (77).

Waterproof. Waterproofing barrier systems may be either hot- or cold-applied. The hot-applied generally involve a bituminous material such as asphalt used in conjunction with a reinforcing fabric such as roofing felt, cotton, or glass cloth. Cold-applied can be bituminous or elastomeric materials either in liquid or sheet form, with or without fabric reinforcement. Liquid elastomeric treatments include neoprene, polyurethanes, and blends of these or epoxies with bituminous materials. Among the commonly used precured elastomeric sheet materials are neoprene, polyisobutylene, EPDM rubber, and plasticized PVC. Polyethylene and PVC films and nonwoven plastic or glass fabric coated with bituminous materials also find use (78). Because these

treatments can trap moisture in a wall, the proper selection and application of treatment materials depend on the nature of the concrete structure, such as whether it is made up of exterior or interior walls above or below grade, floors, or roofs (76).

Water-Repellent. Three techniques used for water repellency are modification of cement by the addition of waterproofers, use of repellent additives to the concrete mix, and surface treatment of concrete structures with repellents. The modification of portland cement by intergrinding with stearate salts or other water-repellent material can reduce the water permeability of mortar. Considerable controversy exists, however, as to whether these cements produce concrete that is superior to carefully mixed concrete without such additives (79).

Admixtures are sometimes used to reduce permeability of concrete (80–82). These include pore-filling materials such as chalk, Fuller's earth, or talc; water repellents such as mineral oil, asphalt, or wax emulsions; organic polymers (acrylic latexes, epoxies); and salts of fatty acids, especially stearates.

The third and perhaps most important class of water repellents consists of materials applied to the surface of concrete for above-grade structures or others where water pressure on the concrete is small. This includes damp-proofing in which treatments cannot be subjected to continuous or even intermittent hydrostatic pressure (83). Repellents that may be used are oils, waxes, soaps, resins, and silicon-based systems (84).

Oils provide short-term benefits on surfaces that are not damaged by the resulting discoloration. For instance, bridge decks may be treated with linseed oil to prevent spalling. Waxes are highly resistant to water, filling pores as well as producing water repellency. Their disadvantages are a tendency to change the appearance of masonry, migration to the surface at high temperatures, and poor abrasion and weathering resistance. Among soaps, the most commonly used are organic solvent solutions of calcium and aluminum stearate.

A wide variety of synthetic polymer resins can be used as water repellents for concrete. These include acrylates, epoxies, chlorinated rubber, polyvinylbutyral, urethanes, elastomeric silicones, and fluoropolymers. They are generally applied from organic solvents or, in some cases, aqueous dispersions or latexes. Some have the disadvantage of inducing gloss and discoloration in the treated masonry.

Concrete can also be made water-repellent by the polymerization of vinyl monomers on the surface (85). Polymerization can be initiated with peroxides, and polyfunctional methacrylates can be used as crosslinking agents. These treatments have a tendency to produce changes in color and gloss.

Siloxanes, alkoxysiloxanes, and alkoxysilanes have in recent years become the most important classes of materials used for water-repellent treatment of masonry, in applications where durability and minimal effect on substrate appearance are important. They achieve these characteristics by penetrating deeply into the masonry (up to 5–10 mm, depending on the substrate and treatment system), hydrophobizing pore surfaces while leaving the surface structure and resultant optical properties largely unaltered. Materials bearing silanol or latent silanol functionality (such as alkoxysilanes, which yield silanols by hydrolysis) can in principle bond covalently to masonry materials through condensation reactions with metal hydroxyl moieties. Silanol self-condensation also leads to cross-linking of the treatment; this combination of reactivity modes, catalyzed by the latent basicity of many masonry types, leads to outstanding durability of water resistance.

Traditionally, materials choice in this area was dominated by solvent-based silanes, siloxane resins or resin–silane combinations (86), and the water-based alkali metal methylsiliconates. The former materials are falling out of favor as a result of a new emphasis on volatile organic compounds (VOC) and solvent reduction, while disadvantages of the latter are their extremely high pH and corrosive nature. In more recent years, aqueous emulsion-based systems have begun to appear which are replacing the older technology.

Two different approaches have been taken toward waterborne alkoxysiloxane and alkoxysilane waterproofing materials. The recent patent literature contains many references to emulsions of alkoxysilanes and silane–resin mixtures (87) that claim stability to hydrolysis until the emulsions are applied and allowed to dry on the masonry, at which point hydrolysis and condensation are underway. Alkoxysilane stability within the emulsion is obtained through judicious choice of emulsifiers and the use of buffering agents to minimize pH drift. Excellent performance in static water holdout tests may be obtained with these systems, although initial water beading by the cured coatings can be somewhat compromised by the presence of surfactant.

A different approach to alkoxysilane or alkoxysiloxane waterproofing agents for masonry is represented by the use of self-emulsifying liquid concentrates (88) intended to be diluted into water and applied within 24–48 h of mixing. These materials owe their ready water-dispersibility to incorporation of low levels of amine functionality protonated by acetic acid; upon drying, the alkoxysilanes hydrolyze and condense, and the acetic acid is lost by volatilization, leaving no residual surfactant behind.

Another class of water-based materials that has recently (ca 1997) begun to see use in masonry water repellency treatments is silicone elastomer latex (89), which can deliver a water-permeable silicone rubber film. These latex elastomers are ideal as water repellents for substrates that contain very large pores, such as concrete block. In addition, the elastomer can bridge minor cracks, and will expand and contract with the substrate.

Oil- and water-resistant treatments have been reported which involve blends of fluoropolymers and alkoxysilane or alkoxysiloxane materials, delivered as 100% actives (90) or from solvent or emulsion (91), fluorinated alkoxysiloxane emulsions (92), and self-emulsifying fluorinated alkoxysilane-based concentrates (93).

Test Methods. Tests for water repellency of concrete and masonry have been described by several authors (72,94). The National Bureau of Standards (NBS) evaluated fifty-five clear water repellents for masonry surfaces (95). Performance tests included water absorption, water-vapor transmission, resistance to efflorescence, change in appearance, and durability to accelerated and outdoor weathering. More recently, the National Cooperative Highway Research Program conducted a study of several organic and silicone water-repellent treatments. The results are communicated in Report 244 (96). The testing procedure used in this study is now a popular method for evaluating water-repellent treatments on concrete.

Wood

Wood is subject to water infiltration by both liquid and vapor. As the moisture content increases, the wood will swell until it reaches its maximum dimension at its fiber-saturation point (about 30% moisture). Variation in the bound water content between zero and 30% will allow the wood to shrink and swell. Rapid dimensional changes resulting from changes in the level of bound water cause the wood to crack and split. These cracks will then allow moisture to absorb easily and quickly into the wood. At moisture content levels above the fiber-saturation point, moisture will be present as free water, which in turn promotes the rate of wood decay.

After-market water repellents, ie, those applied to wood structures after they are built, are widely sold in the consumer market. Materials used in formulating these water repellents are proprietary, but they can be organic resins, waxes, metal stearates, and acrylics. Three components are generally used for a robust water repellent. These include a water-repellent material that penetrates into the wood pores, a material that remains on the surface to give the visual effect of water beading, and a material that reduces the rate of water-vapor transmission into and out of the wood to improve dimensional stability. To penetrate the wood with a solvent or water-based material, a molecular weight of about 1000 or less is generally needed. Water-based short-chain siloxanes which penetrate and bond with the surface are especially effective. Paraffin waxes are typically used to provide surface-water beading. Because it is very difficult to ensure total moisture vapor impermeability, it can be beneficial to use blends of water vapor-permeable materials and water vapor-occlusive materials in order to allow moisture that has entered the wood to exist. Most manufacturers of paints and caulks sell fully formulated water repellents for wood.

Test Methods. Few standard text methods are available to test wood water repellents. Two different types of testing are required to test the water repellency and the dimensional stability. A gravimetric water-repellent method is found in ASTM D5401 (53). Dimensional stability can be tested with ASTM D4446 or Federal Specification TT-W-572B, which is termed the swellometer method. Specifications for the swellometer gauges can be found in ASTM D4446 (53).

BIBLIOGRAPHY

"Waterproofing and Water Repellency" in *ECT* 1st ed., Vol. 14, pp. 862–980, by J. D. Reid and R. K. Worner, U.S. Dept. of Agriculture; in *ECT* 2nd ed., Vol. 22, pp. 135–156, by J. D. Reid and W. J. Connick, Jr., U.S. Dept. of Agriculture; "Waterproofing and Water Oil Repellency" in *ECT* 3rd ed., Vol. 24, pp. 442–465, by M. Hayek, E. I. du Pont de Nemours & Co., Inc.

1. J. L. Moillet, ed., *Waterproofing and Water Repellency*, Elsevier Publishing Co., Amsterdam, the Netherlands, 1963.
2. R. E. Johnson and R. H. Dettre, in J. C. Berg, ed., *Wettability*, Marcel Dekker, Inc., New York, 1993, Chapt. 1.
3. R. E. Johnson and R. H. Dettre, *J. Phys. Chem.* **69**, 1507 (1965).
4. H. W. Fox and W. A. Zisman, *J. Colloid Sci.* **7**, 428 (1952).
5. L. A. Girifalco and R. J. Good, *J. Phys. Chem.* **61**, 904 (1957).
6. R. J. Good, L. A. Girifalco, and G. Kraus, *J. Phys. Chem.* **62**, 1418 (1958).
7. R. J. Good and L. A. Girifalco, *J. Phys. Chem.* **64**, 561 (1960).
8. S. Wu, *J. Polym. Sci. Part C* **34**, 19 (1971).
9. S. Wu, *J. Adhes.* **5**, 39 (1973).
10. M. D. Vobanac and J. C. Berg, *J. Adhesion Sci. Tech.* **4**, 255 (1990).
11. A. B. D. Cassie, *Disc Favalay Soc.* **3**, 239 (1948).
12. F. M. Fowkes, *J. Phys. Chem.* **57**, 98 (1953).
13. F. W. Minor, A. M. Schwartz, E. A. Wulkow, and L. C. Buckles, *Text. Res. J.* **29**, 931 (1959).
14. *AATCC Technical Manual*, American Association of Textile Chemists and Colorists, Research Triangle Park, N.C., 1996.
15. G. R. Lomax, *J. Coated Fabrics* **15**, 40 (1985).
16. M. Van Roey, *J. Coated Fabrics* **21**, 20 (1991).
17. *Textile Chemists and Colorist Buyers Guide*, American Association of Textile Chemists & Colorists, Research Triangle Park, N.C., 1995, pp. 149, 201, 221, 253.
18. W. A. Zisman in R. F. Gould, ed., *Contact Angle, Wettability, and Adhesion*, No. 43 in *Advances in Chemistry*, American Chemical Society, New York, 1964, pp. 1–51.
19. F. Schulman and W. A. Zisman, *J. Am. Chem. Soc.* **74**, 2123 (1952).
20. F. Schulman and W. A. Zisman, *J. Colloid Sci.* **7**, 465 (1952).
21. A. H. Ellison, H. W. Fox, and W. A. Zisman, *J. Phys. Chem.* **57**, 622 (1953).
22. E. F. Hare, E. G. Shafrin, and W. A. Zisman, *J. Phys. Chem.* **58**, 236 (1954).
23. E. G. Shafrin and W. A. Zisman, *J. Phys. Chem.* **64**, 519 (1960).
24. M. K. Bennett and W. A. Zisman, *J. Phys. Chem.* **67**, 1534 (1963).
25. R. E. Johnson, Jr., and R. H. Dettre, in E. Matijevic, ed., *Surface and Colloid Science*, Vol. 2, Wiley-Interscience, New York, 1969, pp. 85–153.
26. R. E. Johnson, Jr., and R. H. Dettre, *J. Colloid Interface Sci.* **21**, 610 (1966).
27. P. B. Roth, *Am. Dyest. Rep.* **60**(7), 34 (1971).
28. E. J. Grajeck and W. H. Peterson, *Text. Res. J.* **32**, 320 (1962).
29. J. F. Colvert, *Tappi* **59**, 129 (1976).
30. J. M. May, *Am. Dyest. Rep.* **58**(20), 15 (1969).
31. J. L. Moillet, ed., *Waterproofing and Water-Repellency*, Elsevier Publishing Co., Amsterdam, the Netherlands, 1963, pp. 52–62.
32. H. A. Schuyten, J. D. Reid, J. W. Weaver, and J. G. Frick, Jr., *Text. Res. J.* **18**, 396 (1948).
33. R. D. Deanin and D. C. Patel, *Text. Res. J.* **40**, 970 (1970).
34. **U.S. Pat. 4,063,885** (Dec. 20, 1977), T. Mares, J. C. Arthur, Jr., and J. A. Harris (to United States of America).
35. N. Nishide and H. Shimizu, *Text. Res. J.* **45**, 591 (1975).
36. M. Sotton and O. Demuth, *Aachener Textiltagung* **114**, 223 (1994).
37. O. C. M. Dorsett, *Text. Manuf.*, 112 (1970).
38. C. A. Davis, *Am. Dyest. Rep.* **56**, 555 (1967).
39. Ref. 37, pp. 7–9.
40. *INDA Standard Test Methods*, Association of the Nonwoven Fabrics Industry, Cary, N.C., 1993.
41. *3M Water Repellency Test II, Water/Alcohol Drop Test*, 3M Test Methods, 3M, St. Paul, Minn., 1992.
42. S. Smith and P. O. Sherman, *Text. Chem. Color.* **1**, 20 (1969).
43. C. J. Bierbrauer, K. D. Goebel, and D. P. Landucci, *Am. Dyest. Rep.* **68**(6), 19 (1979).
44. T. C. Thorstensen, *J. Am. Leather Chem. Assoc.* **73**(5), 196 (1978).
45. F. P. Luvisi, W. J. Hopkins, E. M. Filachione, and J. Naghski, *J. Am. Leather Chem. Assoc.* **64**(12), 614 (1969).
46. **U.S. Pat. 2,834,393** (Apr. 28, 1959), J. W. Gilkey (to Dow Corning Corp.).
47. **U.S. Pat. 3,002,949** (Oct. 3, 1961), S. Nitzsche and E. Pirson (to Wacker Chemie GmbH).
48. R. J. Heit, *J. Am. Leather Chem. Assoc.* **64**(11), 553 (1969).
49. N. V. Vakhrameeva and co-workers, *Kozh. Obuv. Promst.* **23**(6), 43 (1981).
50. K. M. Zurabjan, V. V. Nepommina, and L. V. Slobodskich, *Kozarstvi* **30**(10), 296 (1980).
51. *3M Abrasion Test I*, 3M Test Methods, 3M Co., St. Paul, Minn., 1991.
52. *3M Dynamic Water Resistance Test I (Bally Penetrometer Test)*, 3M Test Methods, 3M, St. Paul, Minn., 1991.
53. *1995 Annual Book of ASTM Standards*, American Society for Testing and Materials, Philadelphia, Pa., 1995.
54. W. F. Reynolds, ed., *The Sizing of Paper*, Tappi Press, Atlanta, Ga., 1989.
55. J. W. Swanson, ed., *Internal Sizing of Paper and Paperboard*, Monograph No. 33, Tappi Press, Atlanta, Ga., 1971.
56. R. W. Davison, *Internal Sizing*, R. W. Hagemeyer, D. W. Manson, and M. J. Kocurek, eds., Pulp and Paper Manufacturing Series, Vol. 6, 1992, Chapt. 3, pp. 39–64.
57. D. Eklund and T. Lindstrom, *Paper Chemistry, An Introduction*, DT Paper Science Publications, Grankulla, Finland, 1991, Chapt. 8, p. 192.
58. J. J. Keavney and R. J. Kulick, in J. P. Casey, ed., *Pulp and Paper, Chemistry and Chemical Technology*, 3rd ed., Vol. 3, Wiley-Interscience, 1981, p. 1547.
59. J. Farewell, ed., *Commercially Available Chemical Agents for Paper and Paperboard Manufacture*, 4th ed., Tappi Press, Atlanta, Ga., 1990.
60. R. M. Husband, ed., *Survey of Paper Additives*, 2nd ed., H&H Consulting Group, Trumbull, Conn., 1991.
61. Ref. 4, p. 134.
62. C. A. Genge, *Anal. Chem.* **31**, 1750 (1959).
63. S. M. Ehrhardt, *1990 Papermakers Conference Proceedings*, Tappi Press, Atlanta, Ga., 1990, p. 171.
64. B. Podd, *World Pulp Paper Technol.*, 243 (1990).
65. S. M. Ehrhardt and J. C. Gast, *1988 Papermakers Conference Proceedings*, Tappi Press, Atlanta, Ga., 1988, p. 181.
66. Ref. 54, p. 63.
67. M. L. Cushing, in Ref. 58, p. 1667.
68. *Code of Federal Regulations*, Title 21, Pts. 174–178, U.S. Government Printing Office, Washington, D.C., 1993.
69. Ref. 54, p. 103.
70. Ref. 58, p. 1550.
71. "Tappi Test Methods", Tappi Press, Atlanta, Ga., 1994.
72. P. Maslow, *Chemical Materials for Construction*, Structures Publishing Co., Farmington, Mich., 1974, pp. 31–34, 433–471.
73. W. H. Taylor, *Concrete Technology and Practice*, McGraw-Hill Book Co., Inc., New York, 1977.
74. I. Biczok, *Concrete Corrosion and Concrete Protection*, Chemical Publishing Co., Inc., New York, 1967.
75. D. F. Orchard, *Concrete Technology*, 4th ed., Applied Science Publishers, London, 1979.

76. *ACI Manual of Concrete Practice*, Pt. 5, Sect. 575.1R-79, American Concrete Institute, Detroit, Mich., 1993.

77. Ref. 73, p. 234.

78. Ref. 76, pp. 20–24.

79. S. Popovics, *Concrete Materials—Properties, Specifications, and Testing*, 2nd ed., Noyes Publications, Park Ridge, N.J., 1992, p. 175.

80. Ref. 72, pp. 31–34.

81. Ref. 73, pp. 234–235.

82. Ref. 79, pp. 258 ff.

83. Ref. 76, pp. 29–30.

84. Ref. 72, pp. 449–454.

85. **U.S. Pat. 4,910,080** (Mar. 20, 1990), G. W. Frost (to 3M Co.).

86. **U.S. Pat. 4,716,051** (Dec. 29, 1987), K. M. Rodder (to Dynamit Nobel AG); **U.S. Pat. 3,772,065** (Nov. 13, 1973), C. D. Seiler (to Dynamit Nobel AG); **U.S. Pat. 4,717,599** (Jan. 5, 1988), D. F. Merrill (to General Electric Co.).

87. **U.S. Pat. 4,648,904** (Mar. 10, 1987), R. J. DePasquale (to SCM Corp.); **U.S. Pat. 4,877,654** (Oct. 31, 1989), M. E. Wilson (to PCR, Inc.); **U.S. Pat. 4,990,377** (Feb. 5, 1991), M. E. Wilson, (to PCR Group, Inc.); **U.S. Pat. 5,037,873** (Aug. 6, 1991), T. F. Heaton (to PCR Group, Inc.); **U.S. Pat. 5,226,954** (July 13, 1993), T. Suzuki (to Tokyo Ink Mfg., Ltd.); **U.S. Pat. 5,393,330** (Feb. 28, 1995), M. J. Chen (to OSi Specialties, Inc.); **U.S. Pat. 5,091,002** (Feb. 25, 1992), E. Schamberg, G. Koerner, H. Fritsch, M. Grasse, and R. Sucker (to Th. Goldschmidt AG); **U.S. Pat. 5,449,712** (Sept. 12, 1995), M. A. Gierke, C. L. Vidal, and M. E. Wilson (to Thoro System Products, Inc.).

88. **U.S. Pat. 4,661,551** (Apr. 28, 1987), H. Mayer, B. Deubner, P. Iretzberger, R. Muhlhofer, and H. Wilhelm (to Wacker-Chemie GmbH); **U.S. Pat. 5,250,106** (Oct. 5, 1993), M. Roth and R. Bernbacher (to Wacker-Chemie GmbH).

89. D. T. Liles and H. V. Lefler III, *Waterborne, Higher-Solids, and Powder Coatings Symposium*, Feb. 1991.

90. **U.S. Pat. 5,112,393** (May 12, 1992), J. F. Engel and G. E. Boyer (to ProSoCo, Inc.).

91. **U.S. Pat. 4,125,673** (Nov. 14, 1978), M. Roth and H. Gluck (to Wacker-Chemie GmbH).

92. **U.S. Pat. 5,442,011** (Aug. 15, 1995), R. A. Halling (to E. I. du Pont de Nemours and Co., Inc.).

93. **Eur. Pat. Appl.** EP 561,708 (Sept. 22, 1993), L. D. Rich and J. F. Sanders (to 3M Co.).

94. V. M. Malhorta, *Testing Hardened Concrete: Nondestructive Methods*, Iowa State University Press, Ames, Ia., and American Concrete Institute, Detroit, Mich., 1976, pp. 136–137.

95. E. J. Clark, P. G. Cambell, and G. Frohnsdorff, *National Bureau of Standards Technocal Note 833*, National Bureau of Standards, Washington, D.C., Oct. 1975, 79 pp.

96. D. W. Pfeifer and M. J. Scali, *Concrete Sealers for Protection of Bridge Structures*, National Cooperative Highway Research Program Report 244, Transportation Research Board, National Academy of Sciences, Washington, D.C., Dec. 1981.

Richard D. Howells
3M

WATER GLASS.

SeeSILICON COMPOUNDS.

WAXES

Wax usually refers to a substance that is a plastic solid at ambient temperature and that, on being subjected to moderately elevated temperatures, becomes a low viscosity liquid. Because it is plastic, wax usually deforms under pressure without the application of heat. The chemical composition of waxes is complex; all of the products have relatively wide molecular weight profiles, with the functionality ranging from products that contain mainly normal alkanes to those that are mixtures of hydrocarbons and reactive functional species.

For centuries, the honeycomb of bees, ie, beeswax, was the material commonly referred to as wax. Substances having typical wax characteristics have traditionally come from insects, eg, beeswax; from vegetables, eg, carnauba; and from animal, eg, spermaceti, origins (1). Waxes from mineral and synthetic sources have been developed both as substitutes for waxes from traditional sources and for new applications. Waxes from minerals and synthetic sources now surpass waxes from traditional sources in tonnage and commercial importance.

Waxes obtained from natural sources such as vegetables or insects are subject to weather conditions which may severely affect the stability of supply and price and, to a lesser extent, the consistency of the products. Waxes from minerals and synthetic sources are less susceptible to weather conditions, and thus have a more stable supply and price.

Insect and Animal Waxes

Beeswax. White [8012-89-3] and yellow [8006-40-4] beeswax has been known for over 2000 years, especially through its use in the fine arts (2). References to wax prior to the nineteenth century are probably to beeswax. Beeswax is secreted by bees and is used to construct the combs in which bees store their honey. The wax is harvested by removing the honey and melting the comb in boiling water; the melted product is then filtered and cast into cakes. The yellow beeswax cakes can be bleached with oxidizing agents to white beeswax, a product much favored in the cosmetic industry. Imports of beeswax into the United States for the years 1990–1995 are listed in Table 1. Historically, Brazil had long supplied the majority of beeswax to the United States, but now (ca 1997) other countries such as China, Thailand, and Canada supply the majority of the material imported into the United States (3).

Table 1. U.S. Imports of Beeswax, Carnauba Wax, Candelilla Wax, and Montan Wax^a

Year	Beeswax		Carnauba wax		Candelilla wax		Montan wax	
	Quantity ^b	Value ^c	Quantity ^b	Value ^c	Quantity ^b	Value ^c	Quantity ^b	Value ^c
1990	1,404	2.49	2,812	2.28	563	2.27	1,664	2.49
1991	823	2.91	3,070	3.65	357	2.49	988	1.71
1992	741	3.02	3,147	3.27	318	2.55	1,369	1.82
1993	901	2.87	3,561	2.30	637	2.61	3,711	0.96
1994	1,111	3.17	2,793	2.70	478	2.65	2,330	1.46
1995	1,382	3.79	3,055	6.08	560	3.03	2,007	1.73

^a Ref. 3.
^b Quantity is in metric tons.
^c Value is in U.S. \$ per kg (\$ kg).

The composition of beeswax varies, depending on its geographic origin. The major components are esters of C₃₀ and C₃₂ alcohols with C₁ acids, free C₂₅ to C₃₁ carboxylic acids, and C₂₅ to C₃₁ hydrocarbons (4). Beeswax typically has a melting point of 64°C, a penetration (hardness) of 20 dmm at 25°C and 76 dmm at 43.3°C (ASTM D1321), a viscosity of 1470 mm² s at 98.9°C, an acid number of 20, and a saponification number of 84.

The U.S. Food and Drug Administration (FDA) has affirmed the status of beeswax as Generally Recognized as Safe (GRAS) (5) in Title 21 of the Code of Federal Regulations, Section 184.1973 (21 CFR 184.1973). The major use of beeswax is in the cosmetic industry, with smaller amounts used in pharmaceuticals and candle production.

Spermaceti. Spermaceti [8002-23-1] is derived from the head oil of the sperm whale. Owing to the present status of the sperm whale as an endangered species, however, the material is no longer an item of commerce and has been replaced by other natural and synthetic waxes.

Vegetable Waxes

Carnauba. The source of carnauba wax [8015-86-9] is the palm tree, whose wax-producing stands grow almost exclusively in the semiarid northeast section of Brazil. Carnauba wax forms on the fronds of the palm, and is removed by cutting the fronds, drying, and mechanically removing the wax. Impurities are removed from the wax by melting and filtering or centrifuging. Wide fluctuations in price and availability have caused markets served by carnauba wax to seek replacements. Whereas there is no other single wax which combines all the properties of carnauba, suitable substitutes are available for most applications. Imports of carnauba wax into the United States for the years 1990–1995 are listed in Table 1. Brazil supplies over 80% of the carnauba wax imported into the United States (3).

The major components of carnauba wax are aliphatic and aromatic esters of long-chain alcohols and acids, with smaller amounts of free fatty acids and alcohols, and resins. Carnauba wax is very hard, with a penetration of 2 dmm at 25°C and only 3 dmm at 43.3°C. Carnauba also has one of the higher melting points for the natural waxes at 84°C, with a viscosity of 3960 mm² s at 98.9°C, an acid number of 8, and a saponification number of 80.

The hardness and high melting point, when combined with its ability to disperse pigments such as carbon black, allows Carnauba wax increasing use in the thermal printing inks. Carnauba is also widely used to gel organic solvents and oils, making the wax a valuable component of solvent and oil paste formulations. Carnauba polishes to a high gloss and thus is widely used as a polishing agent for items such as leather, candies, and pills. Other uses include cosmetics and investment casting applications (see COPPER ALLOYS, CAST COPPER ALLOYS). The FDA affirmed Camauba as GRAS for certain food applications in 21 CFR 184.1978.

Candelilla. Candelilla wax [8006-44-8] is harvested from shrubs in the Mexican states of Coahuila and Chihuahua and, to a very small degree, in

the Big Bend region of Texas in the United States (6). The entire mature plant is uprooted and immersed in boiling water acidified with sulfuric acid. The wax floats to the surface and is filtered. Imports of candelilla wax into the United States for the years 1990–1995 are listed in Table 1. Mexico supplies over 90% of the candelilla wax imported into the United States (3).

The major components of candelilla wax are hydrocarbons, esters of long-chain alcohols and acids, long-chain alcohols, sterols, and neutral resins, and long-chain acids. Typically, candelilla wax has a melting point of 70°C, a penetration of 3 dmm at 25°C, an acid number of 14, and a saponification number of 55. Principal markets for candelilla include cosmetics, foods, and pharmaceuticals. The FDA affirmed Candelilla as GRAS for certain food applications in 21 CFR 184.1976.

Japan Wax. Japan wax [8001-39-6] is a fat and is derived from the berries of a small tree native to Japan and China cultivated for its wax. Japan wax is composed of triglycerides, primarily tripalmitin. Japan wax typically has a melting point of 53°C, an acid number of 18, and a saponification number of 217. Principal markets include the formulation of candles, polishes, lubricants, and as an additive to thermoplastic resins. The product has some food-related applications.

Ouricury Wax. Ouricury wax [68917-70-4] is a brown wax obtained from the fronds of a palm tree which grows in Brazil. Ouricury is difficult to harvest, as it does not flake off the frond as does carnauba wax; rather, it must be scraped free. Ouricury is sometimes used as a replacement for carnauba in applications that do not require a light-colored wax.

Rice-Bran Wax. Rice-bran wax [8016-60-2] is extracted from crude rice-bran oil. It can be degummed, the fatty acid content reduced by solvent extraction, and bleached. The wax is primarily composed of esters of lignoceric acid (43 wt %), behenic acid (16 wt %), and C₂₂–C₃ alcohols (28 wt %). Rice-bran wax may be used in some food applications under the regulations described in 21 CFR 172.890.

Jojoba. Jojoba oil [61789-91-1] is obtained from the seeds of the jojoba plant grown in semiarid regions of Costa Rica, Israel, Mexico, and the United States. The oil is made up of ca 80 wt % of esters of eicos-11-enoic and docos-13-enoic acids, and eicos-11-*en*-1-ol, and docos-13-*en*-1-ol, ca 17 wt % of other liquid esters, with the balance being free alcohols, free acids, and steroids. Jojoba oil is used primarily in the formulation of cosmetics. Hydrogenated jojoba oil is a wax used in candles and other low volume specialty applications.

Castor Wax. Castor wax [8001-78-3] is catalytically hydrogenated castor bean oil. The wax has a melting point of 86°C, acid number of 2, saponification number of 179, and an iodine number of 4. Castor wax is used primarily in the formulation of cosmetics. Derivatives of castor wax are used as surfactants and plastics additives.

Bayberry Wax. Bayberry wax [8038-77-5] is removed from the surface of the berry of the bayberry (myrtle) shrub by boiling the berries in water and skimming the wax from the surface of the water. The wax is green and made up primarily of lauric, myristic, and palmitic acid esters. The wax has a melting point of 45°C, an acid number of 15, a saponification number of 220, and an iodine number of 6. The wax has an aromatic odor and is used primarily in the manufacture of candles and other products where the distinctive odor is desirable.

Mineral Waxes

Montan Wax. Montan wax [8002-53-7] is derived by solvent extraction of lignite (qv). The earliest production on a commercial scale was in Germany during the latter half of the nineteenth century, and Germany continues to supply the majority of the world's production of montan wax. Montan wax production at Amsdorf is part of a massive coal-mining operation from a continuous vein and raw material is expected to last for decades. Montan wax is also produced in the United States. Imports of montan wax into the United States for the years 1990–1995 are listed in Table 1. Germany supplies over 80% of the montan wax imported into the United States (3).

The composition of montan wax depends on the material from which it is extracted, but all contain varying amounts of wax, resin, and asphalt. Black montan wax may be further processed to remove the resins and asphalt, and is known as refined montan wax. White montan wax has been reacted with alcohols to form esters. The wax component of montan is a mixture of long-chain (C₂₄–C₃₀) esters (62–68 wt %), long-chain acids (22–26 wt %), and long-chain alcohols, ketones, and hydrocarbons (7–15 wt %). Crude montan wax from Germany typically has a melting point of 80°C, an acid number of 32, and a saponification number of 92.

The largest traditional use for montan waxes was as a component in one-time hot-melt carbon-paper inks. With the decrease in the use of carbon-paper inks, uses for the refined grades have become predominant, mainly in the formulation of polishes and as plastics lubricants. The alcohol ester derivatives may be used as components of articles intended for use with foods as regulated by the FDA in 21 CFR 178.3770.

Peat Waxes. Peat waxes are much like montan waxes in that they contain three main components: a wax fraction, a resin fraction, and an asphalt fraction. The amount of asphalt in the total yield is influenced strongly by the solvent used in the extraction. Montan waxes contain ca 50 wt % more of the wax fraction than peat waxes, and correspondingly lower percentages of the resin and asphalt fractions. The wax fraction in peat wax is chemically similar to that of the wax fraction in montan wax.

Ozokerite and Ceresin Waxes. Ozokerite wax [8001-75-0] was a product of Poland, Austria, and in the former USSR where it was mined. True ozokerite no longer seems to be an article of commerce, and has been replaced with blends of petroleum-derived paraffin and microcrystalline waxes. These blends are designed to meet the specific physical properties required by the application involved.

Ceresin wax [8001-75-0] originally was a refined and bleached ozokerite wax, but now is a paraffin wax of very narrow molecular weight distribution or blend of petroleum waxes.

Petroleum Waxes. Waxes derived from petroleum are hydrocarbons of three types: paraffin [64742-43-4] (clay-treated); semimicrocrystalline or intermediate; and microcrystalline [64742-42-3] (clay-treated). Semimicrocrystalline waxes are not generally marketed as such (7). Others include acid-treated, chemically neutralized, and hydrotreated; and paraffin and hydrocarbon waxes, untreated. The quality and quantity of the wax separated from the crude oil depends on the source of the crude oil and the degree of refining to which it has been subjected prior to wax separation. Petroleum waxes are produced in massive quantities throughout the world. Subject to the wax content in the crude, paraffin and, to a substantially lesser degree, microcrystalline wax are produced in almost all countries of the world that refine crude oil. Production capacity in the United States and imports for the years 1990 to 1995 are listed in Table 2. Canada supplies over 50% of the petroleum wax imported into the United States (3).

Table 2. U.S. Production Capacity and Imports of Petroleum Wax

Year	U.S. Capacity ^a , t	Import Data ^b	
		Quantity, t	Value ^c
1990	585,300	81,306	0.58
1991	551,200	79,176	0.60
1992	624,300	105,297	0.56
1993	614,600	89,330	0.53
1994	634,100	120,599	0.74
1995	643,900	113,800	0.57

^b Ref. 3.

^a Ref. 8.

^c Value in U.S. \$ per kg (\$ kg).

A paraffin wax is a petroleum wax consisting principally of normal alkanes. Microcrystalline wax is a petroleum wax containing substantial proportions of branched and cyclic saturated hydrocarbons, in addition to normal alkanes. Semimicrocrystalline wax contains more branched and cyclic compounds than paraffin wax, but less than microcrystalline. A classification system based on the refractive index of the wax and its congealing point as

determined by ASTM D938 was developed (9).

Paraffin wax is macrocrystalline, brittle, and is composed of 40–90 wt % normal alkanes, with the remainder C_{18} – C_{30} isoalkanes and cycloalkanes. Paraffin wax has little affinity for oil content: fully refined paraffin has less than 1 wt %; crude scale, 1–2 wt %, and slack [64742-61-6], above 2 wt %. Within these classes, the melting point of the wax determines the actual grade, with a range of about 46–71°C. Typical properties of petroleum waxes are listed in Table 3.

Table 3. Typical Properties of Petroleum Waxes

Property	Wax	
	Paraffin	Microcrystalline
flash point, closed cup, °C	204 ^a	260 ^a
viscosity at 98.9°C, mm ² s	4.2–7.4	10.2–25
melting range, °C	46–68	60–93
refractive index at 98.9°C	1.430–1.433	1.435–1.445
number average molecular weight	350–420	600–800
carbon atoms per molecule	20–36	30–75
ductility crystallinity of solid wax	Friable to crystalline	Ductile–plastic to tough–brittle

^a Value is minimum.

The separation of paraffin wax from crude oil occurs during distillation, as shown in Figure 1. The distillate is processed to remove oil to the degree desired through solvent extraction. It is then decolorized, usually by hydrogenation, but percolation through bauxite is also used. Microcrystalline wax is produced either from the residual fraction of crude oil distillation or from crude oil tank bottoms (10). After deasphalting of the residual fraction, heavy lubricating oil is removed by solvent extraction. The degree of solvent extraction is dictated by the economics of the lubrication oil market. The filtrate is crude petrolatum, a dark-colored, unctuous material containing oil and microcrystalline wax. Percentages of each may vary, but are usually about 40 wt % wax and 60 wt % oil. This material is then solvent-extracted for the wax. Because microcrystalline wax has great affinity for oil, the oil content of the wax is 1–4 wt %, depending on the grade of the wax. Unlike paraffin wax, oil is held tightly in the crystal lattice of the microcrystalline wax, and does not migrate to the surface. The microcrystalline waxes obtained from petrolatums are generally known as plastic grades, with penetrations greater than 11 dmm at 25°C.

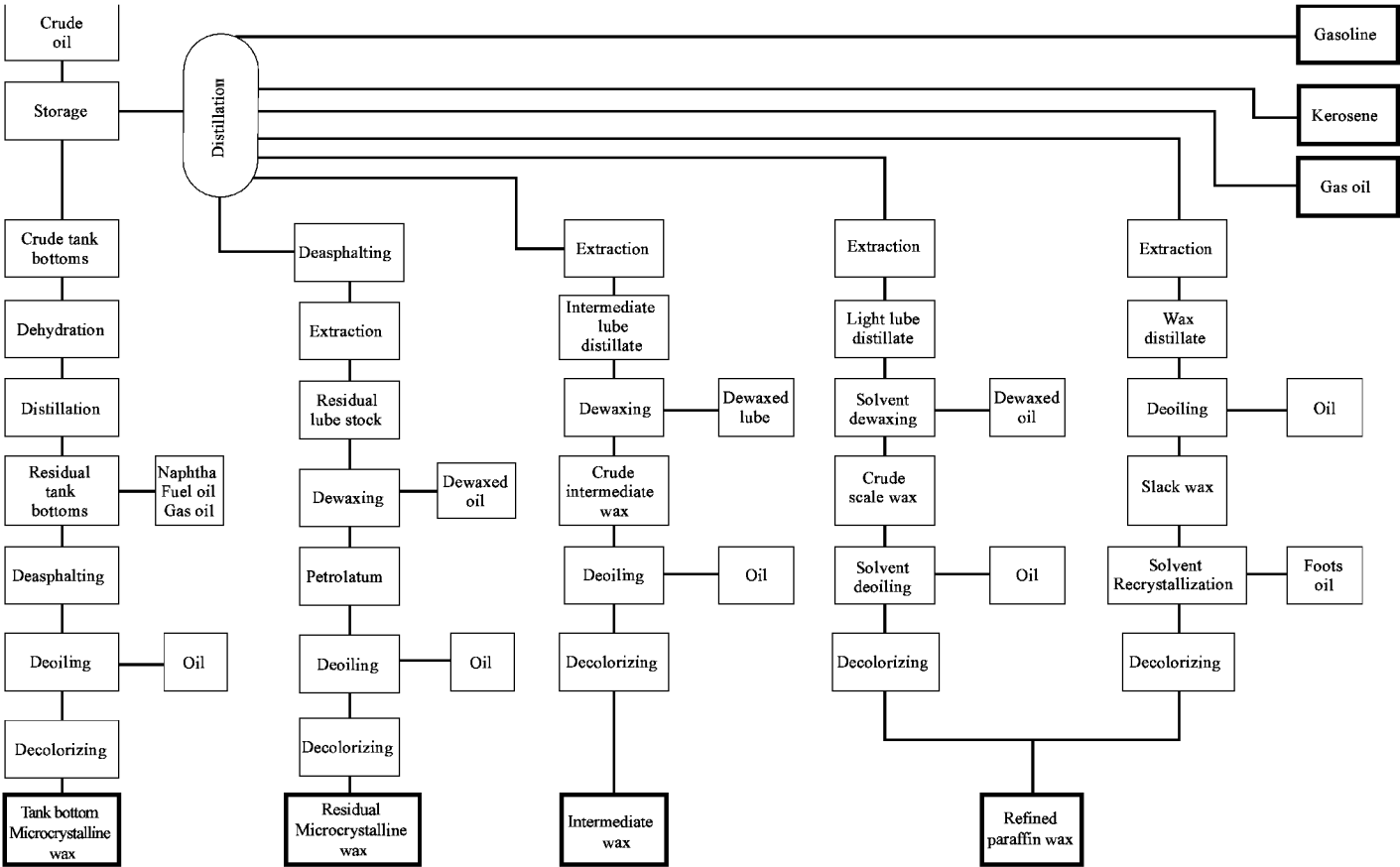


Fig. 1. Refining of petroleum waxes. Courtesy of Baker Petrolite Corporation.

Crude oil contains high molecular weight fractions which are soluble at the high temperatures found in underground formations, but not very soluble at ambient conditions once the crude oil is produced. These high molecular weight fractions precipitate onto the walls and floors of storage tanks, and are known as crude oil tank bottoms. Crude oil tank bottoms are essentially crude oil with very high wax contents and are processed as indicated in Figure 1. The microcrystalline waxes obtained from crude oil tank bottoms are generally known as hard grades, with penetrations less than 11 dmm at 25°C.

The use of refined grades of petroleum waxes in some food applications is regulated by the FDA in 21 CFR 172.886 and 21 CFR 178.3710. The Bundesgesundheitsamt (BGA) of Germany also has specifications for refined petroleum waxes used in food applications. Many other countries reference either the FDA or BGA specifications for their food regulations. Petroleum wax is widely used in chewing gum to modify the properties of the chewing gum base. The wide range of properties available help chewing gum base manufacturers formulate a broad variety of chewing gum, ranging from the traditional hard stick gum to the softer bubble gum. Petroleum wax can also be used as protective coatings for fruits, vegetables, and cheeses. Petroleum wax is outstanding as a cost-effective moisture and gas barrier, and food packaging applications are a major market for refined food-grade petroleum wax. Blends of paraffin and microcrystalline wax are used by themselves or in combination with other additives such as high molecular weight polyethylene and

ethylene vinyl acetate copolymers to improve the performance of paper packaging such as paperboard boxes, paper containers, and flexible packaging. Petroleum waxes are also widely used in other industrial applications. Paraffin waxes are added to rubber during compounding, and exude to the surface during curing, which helps protect the rubber from degradation resulting from ozone. Paraffin and other waxes can be added to plastics, especially poly(vinyl chloride) (PVC) as lubricants. Both paraffin and microcrystalline waxes are widely used to help control the properties of hot-melt adhesives. Dispersions of microcrystalline are added to inks to improve slip and rub properties. Petroleum waxes are used in many consumer applications such as cosmetics, polishes, and candles. Unrefined petroleum waxes are often used in fireplace logs.

Synthetic Waxes

Polyethylene Waxes. Low molecular weight (less than ca 10,000 Mn) polyethylenes [9002-88-4] having waxlike properties are made either by high pressure polymerization or low pressure (Ziegler-type catalysts) polymerization. All the products have the same basic structure, but the processes yield products with distinctly different properties. Some polyethylenes have fairly low densities, owing to branching that occurs during the polymerization. Molecular weight distributions, expressed as the weight average molecular weight divided by the number average molecular weight, or polydispersity, also varies widely among the different processes, as does the range of molecular weights available. Annual production in the United States is estimated at 100,000–140,000 t.

Differences among the processes have a major impact on the use of the products. Products from a particular process or manufacturer may dominate one market, while products from a different process may be preferred in a different application. Major uses include hot-melt adhesives for applications requiring high temperature performance, additives to improve the processing of plastics, slip and rub additives for inks and paints, and cosmetic applications.

Products used in food applications require regulatory approvals. Most polyethylenes are regulated by the FDA under the olefin polymer regulation, 21 CFR 177.1520. This regulation includes a maximum amount of hexane-soluble material with other requirements. The amount of material extracted by hexane is a function of molecular weight and branching. Polyethylenes in the 500–1200-molecular weight range are regulated by the FDA under the synthetic petroleum wax regulation, 21 CFR 172.888. In addition to molecular weight requirements, this regulation includes an absorbance test to verify the suitability of the product for food applications.

Some by-product polyethylene waxes have been recently introduced. The feedstock for these materials are mixtures of low molecular weight polyethylene fractions and solvent, generally hexane, produced in making polyethylene plastic resin. The solvent is stripped from the mixture, and the residual material offered as polyethylene wax. The products generally have a wider molecular weight distribution than the polyethylene waxes synthesized directly, and are offered to markets able to tolerate that characteristic. Some of the by-product polyethylene waxes are distilled under vacuum to obtain a narrower molecular weight distribution.

Several of the polymerization processes allow different functionality to be added to the backbone of the polymer, including copolymers of ethene, propene, hexene, vinyl acetate, and acrylic acid, with waxlike properties. Copolymers of ethene with other olefins provide a method of extending the range of properties available. The addition of other olefins creates a branched polymer, which decreases the melting point and hardness, while increasing viscosity as compared to a linear polyethylene of the same molecular weight distribution. Longer branches created through the addition of hexene show a larger effect than those from propene. Copolymers with vinyl acetate and acrylic acid provide a method of introducing oxygen functionality. These products may be further reacted with metal salts to form ionomers.

In addition to copolymerization, polyethylenes terminated as ketones, alcohols, and carboxylic acids with molecular weights as high as 700 daltons are now available. The products offer the same chemical functionality as common fatty alcohols and acids, but are higher melting and harder. Similar to the fatty alcohols and acids, derivatives such as ethoxylates, esters, and amides also are available as higher melting versions of the fatty derivatives.

Functional polyethylene waxes provide both the physical properties obtained by the high molecular weight polyethylene wax and the chemical properties of an oxidized product, or one derived from a fatty alcohol or acid. The functional groups improve adhesion to polar substrates, compatibility with polar materials, and dispersibility into water. Uses include additives for inks and coatings, pigment dispersions, plastics, cosmetics, toners, and adhesives.

Fischer-Tropsch Waxes. Polymethylene wax [8002-74-2] production is based on the Fischer-Tropsch synthesis, which is basically the polymerization of carbon monoxide under high pressure and over special catalysts to produce hydrocarbons (see FUELS, SYNTHETIC LIQUID FUELS). Distillation is then used to separate the hydrocarbons into different products, including liquid fuels and waxes with melting points ranging from about 45–106°C. Currently the waxes are produced in large volumes in South Africa and Malaysia, with an estimated 12,000–14,000 t consumed in the United States in 1994. Uses are similar to those for polyethylene waxes, including hot-melt adhesives and additives for inks and coatings.

Chemically Modified Waxes. Hydrocarbon waxes of the microcrystalline, polyethylene, and polymethylene classes are chemically modified to meet specific market needs. In the vast majority of cases, the first step is air oxidation of the wax with or without catalysts (11). The product has an acid number usually no higher than 30 and a saponification number usually no lower than 25. An alternative step is the reaction of the wax with a polycarboxylic acid, eg, maleic, at high temperature (12). Through its carboxyl groups, the oxidized wax can be further modified in such reactions as saponification or esterification. Oxidized wax is easily emulsified in water through the use of surfactants or simple soaps, and is widely used in many coating and polish applications.

Substituted Amide Waxes. The product of fatty acid amidation has unique waxlike properties (13). Probably the most widely produced material is *N,N'*-distearylthylenediamine [110-30-5], which has a melting point of ca 140°C, an acid number of ca 7, and a low melt viscosity. Because of its unusually high melting point and unique functionality, it is used in additive quantities to raise the apparent melting point of thermoplastic resins and asphalts, as an internal–external lubricant in the compounding of a variety of thermoplastic resins, and as a processing aid for elastomers.

Polymerized α -Olefins. Some polymers of higher α -olefins, eg, $C > 20$, have waxlike properties and are sold as synthetic waxes. The polymerization process yields highly branched materials, with broad molecular weight distributions. Properties of the individual products are highly dependent on the α -olefin monomers and polymerization conditions. Melting points for the products range from 54°C to 74°C, with number average molecular weights ca 2600–2800, and penetrations at 25°C of 5–12 dmm. The unique structure makes these products very effective when used in additive amounts to modify the properties of paraffin wax, primarily for use in candles. The products can increase the hardness and opacity of the paraffin, without increasing the cloud point or viscosity. Other uses include mold release for polyurethane foams, additives for casting wax, and additives for leather treating.

Analytical Techniques

Most waxes are complex mixtures of molecules with different carbon lengths, structures, and functionality. Attempts to measure the exact chemical composition are extremely difficult, even for the vegetable waxes, which are based on a relatively few number of basic molecules. Products such as oxidized microcrystalline wax not only have a mixture of hydrocarbon lengths and types as starting materials, but also add complexity through the introduction of various types of carboxylic functionality onto those hydrocarbons during the oxidation process.

Because of the difficulty in analysis of chemical composition, most of the routine test procedures on waxes are for the measurement of the physical properties of the waxes and are used to compare the properties of waxes within a class. Some properties, such as acid number or saponification number, give insight into the chemical functionality of the product, and are widely used for products which contain carboxyl groups such as vegetable, montan, and oxidized waxes. Increasingly, instrumental methods such as gas chromatography (gc), gel permeation (also known as size exclusion) chromatography (gpc), refractive index (ri), differential scanning calorimetry (dsc), infrared spectroscopy (ir), and nuclear magnetic resonance (nmr) are being used to further characterize the products. Properties such as molecular weight distribution, degree of branching, degree of crystallinity, and functionality can be readily measured with these techniques.

Melting and Congealing Points. Selection of the proper melting point method depends upon the characteristics of the wax. Drop melting point (ASTM D127) is suitable for amorphous waxes, eg, microcrystallines, but is not reliable for higher viscosity synthetic waxes, for which ring-and-ball softening point (ASTM D36) should be used. Waxes whose time–temperature cooling curves exhibit plateaus, eg, paraffin wax, may be evaluated by ASTM

D87. Open or closed capillary tubes are used to measure the melting point of many of the natural waxes. The congealing point (ASTM D938) is the temperature at which a melted wax ceases to flow, and is more consistent than melting points for some waxes.

Hardness (Penetration). The standard test for the hardness of waxes in industry is the penetration test (ASTM D1321). This test measures the depth in tenths of a millimeter that a needle of a certain configuration under a given weight penetrates the surface of a wax at a given temperature. A series of penetrations measured at different temperatures, rather than at a single temperature, is preferred.

Color. On solidification of a wax and depending on factors such as the rate of cooling, the amount of occluded air, and surface finish, the color of solidified samples of the same wax may be different. For this reason, the color of most waxes is judged only while molten, although some commercial standards for certain waxes, eg, carnauba, are based on the color of the solid wax. The accurate measurement of color in light-colored, ie, amber to off-white to white, waxes is difficult but very important because of the additional processing costs required to achieve the light color. The two most widely used color standards providing numerical measurement are ASTM D1500, which is used to measure dark-brown to off-white color, and ASTM D156, which is used to measure off-white to pure white.

Oil Content. The production of petroleum waxes involves the removal of oil; therefore, the oil content (actually the percentage of oil and low molecular weight fractions) is one indication of the quality of the wax. Oil content is determined (ASTM D721) as that percentage of the wax soluble in methyl ethyl ketone at 31.7 °C.

Viscosity. Although traditionally of little importance in the evaluation of vegetable and insect waxes, viscosity is an important test for mineral and synthetic waxes. One of the most frequently used tests, ASTM D88, is used to measure the time in seconds required for a specified quantity of wax at a specified temperature to flow by gravity through an orifice of specified dimensions. This viscosity is expressed in Saybolt Universal Seconds (SUS) at the temperature of the test. The SI unit for kinematic viscosity is mm² s (cSt).

Acid Number. The acid number (ASTM D1386) is the milligrams of potassium hydroxide necessary to neutralize one gram of wax, and indicates the amount of free carboxylic acid present. The test is widely used for vegetable and insect waxes, and synthetic waxes containing carboxylic acid groups.

Saponification Number. The saponification number (ASTM D1387) is the milligrams of potassium hydroxide which react with one gram of wax under elevated temperatures, and indicates the amount of free carboxylic acid plus any ester materials which may be saponified. Both the acid number and saponification numbers are generally provided to give an indication of the free carboxylic acid and ester content of vegetable and insect waxes, and synthetic waxes containing carboxylic acids and or esters.

Differential Scanning Calorimetry (dsc). The dsc has become widely used to characterize waxes. Under controlled heating and cooling rates, the amount of energy consumed or released is measured. Curves of heat flow vs temperature provide insight into the thermal characteristics of a wax, including crystalline transitions such as solid-to-solid, solid-to-liquid, and liquid-to-solid. Common values obtained from the curves include the initial and ending temperatures for heat flow, and heat of fusion, expressed as joules per gram.

Gas Chromatography (gc). Gas chromatography has been used for many years, especially on the relatively simple structures of vegetable and insect waxes. Use of the gc for petroleum and synthetic waxes was limited by the maximum carbon number which could be eluted, and the number of isomers for each carbon number. Improvements in technology have allowed wider use of this technique, with columns and equipment available which can resolve carbon numbers up to C₁₀₀. Good resolution can be obtained on products with generally only one type of structure, eg, paraffins with a high preponderance of primary alkanes. Products such as microcrystalline wax which contain several different branched isomers for each carbon number, plus some cyclic compounds, cannot be completely resolved, although useful information can still be obtained.

Gel Permeation Chromatography (gpc). The gpc (also known as size exclusion chromatography) is widely used to measure the molecular weight distribution for synthetic polyethylene waxes. Whereas gpc can not match the resolution available in gc techniques, useful information regarding the molecular weight, and molecular weight distribution can be obtained for products with molecular weights too high for gas chromatography. The molecular weight is normally reported using the number average, M_n , or the weight average, M_w . The ratio of the weight average to the number average is known as the polydispersity, P_d .

Infrared Spectroscopy (ir). Infrared curves are used to identify the chemical functionality of waxes. Petroleum waxes with only hydrocarbon functionality show slight differences based on crystallinity, while vegetable and insect waxes contain hydrocarbons, carboxylic acids, alcohols, and esters. The ir curves are typically used in combination with other analytical methods such as dsc or gc gpc to characterize waxes.

Nuclear Magnetic Resonance (nmr). The nmr analysis has been used in the polymer industry for some time to measure properties such as amount and type of branching, polymerized ethylene oxide content, and hydroxyl content. The same techniques are applicable to waxes, and are used for both characterization and quality control.

BIBLIOGRAPHY

"Waxes" in *ECT* 1st ed., Vol. 15, pp. 1–17, by C. J. Marsel, New York University; in *ECT* 2nd ed., Vol. 22, pp. 156–173, by E. S. McLoud, Consultant; in *ECT* 3rd ed., Vol. 24, pp. 466–481, by C. S. Letcher, Petrolite Corporation.

1. P. E. Kolattukudy, *Chemistry and Biochemistry of Natural Waxes*, Elsevier Scientific Publishing Co., Amsterdam, the Netherlands, 1976.
2. R. White, *Stud. Conserv.* **23**, 57 (1978).
3. *U.S. Imports of Merchandise*, U.S. Bureau of Census, Washington, D.C., 1990–1995.
4. A. H. Warth, *The Chemistry and Technology of Waxes*, Reinhold Publishing Corporation, New York, 1947, p. 49.
5. *Code of Federal Regulations*, CFR 21, Part 184.1973, U.S. Government Printing Office, Washington, D.C., published annually.
6. *Monograph on Candelilla Wax*, PB 287-762, National Technical Information Service, U.S. Department of Commerce, Washington, D.C., Aug. 10, 1978.
7. W. M. Mazee, *Modern Petroleum Technology*, 4th ed., Applied Science Publishers, Ltd., Barking, U.K., pp. 782–803.
8. *1990–1995 Lubricating Oil and Wax Capacities of U.S. and Canadian Refineries*, National Petroleum Refiners Association, Washington, D.C., published annually.
9. *Petroleum Waxes—Characterization, Performance, and Additives*, STAP No. 2, Technical Association of the Pulp and Paper Industry, Atlanta, Ga., 1963, p. 16.
10. F. J. Ludwig, *Anal. Chem.* **37**, 1732 (Dec. 1965).
11. **U.S. Pat. 3,060,163** (Oct. 23, 1962), M. Erchak, Jr. (to Allied Chemical Corp.).
12. **U.S. Pat. 2,766,214** (Oct. 9, 1956), M. Erchak, Jr. (to Allied Chemical Corp.).
13. E. S. Pattison, ed., *Fatty Acids and Their Industrial Applications*, Marcel Dekker, Inc., New York, 1968, p. 85.

William P. Cottom
Baker Petrolite Corporation

WEED KILLERS.

See HERBICIDES.

WOOD

Wood is an important natural resource, one of the few that are renewable. It is prevalent in our everyday lives and the economy; in wood-frame houses and furniture; newspapers, books, and magazines; bridges and railroad ties; fence posts and utility poles; fuelwood; textile fabrics; and organic chemicals. Wood and wood products are also a store for carbon, thus, helping to minimize carbon dioxide in the atmosphere.

Wood supplies the solid raw material for products, such as lumber, plywood, and wood pallets, and the fiber for paper, paperboard, fiberboard panels, rayon, and acetate (see also WOOD; PAPER; PULP). Many wood products can be recovered for reuse or recycling, thus extending our wood supply into the future. A large and increasing portion of the paper and paperboard products that is used, such as newspapers, magazines, and corrugated containers, are recovered for recycling: 45° of the total new supply in 1995 (1). In addition, wood residues from lumber, plywood, and pulp mills are recovered and used to make new fiber products or burned to generate energy.

Production and consumption of wood products and residues are measured in various units, based on common usage and their metric equivalents (2–4). Pulpwood logs and fuelwood are commonly measured in cords. A cord refers to a stacked pile of wood, with outside dimensions of 4 by 4 by 8 ft (1.22 by 1.22 by 2.44 m) and a volume of 128 ft³ (3.62 m³). The weight of a cord depends on density of wood and bark and on moisture content. In the United States, it can range from 1.3 to 1.7 short tons (1.2 to 1.5 metric tons), air dried.

Sawn lumber is commonly measured in board feet. A board foot of lumber has a nominal dimension of 1 ft by 1 ft by 1 in. (30.5 by 30.5 by 2.5 cm) and volume of 1 12 ft³ (2360 cm³ 0.0024 m³). The actual board foot dimension and volume is often smaller, however. Paper and paperboard products are commonly measured in short tons (2,000 lb) or metric tons (2,204.6 lb).

Roundwood equivalent is also sometimes used. It refers to the volume of logs or other round products required to produce a given quantity of lumber, plywood, wood pulp, and other wood products (2). Roundwood equivalents may be used in assessing the overall wood resource supply and demand or in comparing wood products measured in different units.

In recent years, lumber production has accounted for close to 40° of all roundwood used in the United States (4). Pulp, paper, and board products have accounted for close to 30° of total roundwood used. Between 1965 and 1994, annual production of lumber (in board feet) increased by 26° . In contrast, production of paper and board (in tons) more than doubled (4).

The following shows production levels for selected major commodities and years (4). Lumber is given in nominal volume terms (1,000 board feet 2.36 m³).

	1970	1980	1990	1994
Lumber (million cubic meters)	81.7	83.3	114.5	109.5
Plywood (million cubic meters)	14.1	14.6	19.5	18.5
Paper (shipments; million metric tons)	22.9	28.6	35.7	39.5
Paperboard (million metric tons)	23.2	28.4	35.7	41.5

Structure

The anatomical structure of wood affects strength properties, appearance, resistance to penetration by water and chemicals, resistance to decay, pulp quality, and the chemical reactivity of wood (5). To use wood most effectively requires a knowledge of not only the amounts of various substances that make up wood, but also how those substances are distributed in the cell walls.

Woods are either hardwoods or softwoods. Hardwood trees (angiosperms, ie, plants with covered seeds) generally have broad leaves; are deciduous in the temperate regions of the world; and are porous, ie, they contain vessel elements. Softwood trees (conifers or gymnosperms, ie, plants with naked seeds) are cone bearing; generally have scalelike or needlelike leaves; and are nonporous, ie, they do not contain vessel elements. The terms *hardwood* and *softwood* have no direct relation to the hardness or softness of the wood. In fact, hardwood trees such as cottonwood, aspen, and balsa have softer wood than the western white pines and true firs; certain softwoods, such as longleaf pine and Douglas-fir, produce wood that is much harder than that of basswood or yellow-poplar.

Many mechanical properties of wood, such as bending and crushing strength and hardness, depend upon the density of wood; denser woods are generally stronger (6). Wood density is determined largely by the relative thickness of the cell wall and by the proportions of thick-walled and thin-walled cells present.

The cells that make up the structural elements of wood are of various sizes and shapes and are firmly bonded together. Dry wood cells may be empty or partly filled with deposits such as gums, resins, or other extraneous substances. Long and pointed cells, known as fibers or tracheids, vary greatly in length within a tree and from species to species. Hardwood fibers are ~1 mm long, and softwood fibers are ~3 to 8 mm.

Just under the bark of a tree is a thin layer of cells, not visible to the naked eye, called the cambium. Here, cells divide and eventually differentiate to form bark tissue outside of the cambium and wood or xylem tissue inside of the cambium. This newly formed wood on the inside contains many living cells and conducts sap upward in the tree, and hence, is called sapwood. Eventually, the inner sapwood cells become inactive and are transformed into heartwood. This transformation is often accompanied by the formation of extractives that darken the wood, make it less porous, and sometimes provide more resistance to decay.

Because of the great structural variations in wood (7), there are many possibilities for selecting a species for a specific purpose. Some species, like spruce, combine light weight with relatively high values for stiffness and bending strength. Very heavy woods, like lignumvitae, are extremely hard and resistant to abrasion. A very light wood, like balsa, has high thermal insulation value. Hickory has extremely high shock resistance. Mahogany has excellent dimensional stability.

Composition

Wood is a complex polymeric structure consisting of lignin (qv) and carbohydrates (qv) [cellulose (qv) and hemicelluloses], which form the visible

lignocellulosic structure of wood (8–10). Also present, but not contributing to wood structure, are minor amounts of other organic chemicals and minerals. The organic chemicals are diverse and can be removed from the wood with various solvents. The minerals constitute the ash residue remaining after ignition at a high temperature.

Wood species cannot readily be determined by chemical analysis because composition is affected by many variables, including geographical location, soil and weather conditions, and location of the wood within a given tree. Some generalizations are possible to distinguish hardwood and softwood composition, eg, the average lignin content of softwood is slightly higher than that of hardwood. If the minerals and small amounts of nitrogen and sulfur (0.1–0.2%) are ignored, the elementary composition of dry wood averages 50% carbon, 6% hydrogen, and 44% oxygen.

Lignin. Lignin is an amorphous, insoluble organic polymer and is very difficult if not impossible to isolate in a natural state. Molecular weights of isolated lignins range from the low thousands to as high as 50,000. The basic chemical structural unit is a methoxy-substituted propylphenol moiety, bonded in an irregular pattern of ether and carbon–carbon linkages. Lignin comprises 18–30% by weight of the dry wood, most of it concentrated in the compound middle lamella and the layered cell wall. It imparts a woody, rigid structure to the cell walls and distinguishes wood from other fibrous plant materials of lesser lignin content. Quantitative analysis usually involves removing the carbohydrate material by acid hydrolysis and filtering and weighing the insoluble residue (referred to as Klason lignin).

Carbohydrates. Carbohydrates are the principal components of the cell wall, comprising 65–75% by weight of the dry wood. Total hydrolysis yields simple sugars, primarily glucose and xylose in hardwoods and glucose and mannose in softwoods. Minor amounts of galactose, arabinose, and rhamnose are present.

Cellulose is the main component of the wood cell wall, typically 40–50% by weight of the dry wood. Pure cellulose is a polymer of glucose residues joined by 1,4-β-glucosidic bonds. The degree of polymerization (DP) is variable and may range from 700 to 10,000 DP or more. Wood cellulose is more resistant to dilute acid hydrolysis than hemicellulose. X-ray diffraction indicates a partial crystalline structure for wood cellulose. The crystalline regions are more difficult to hydrolyze than the amorphous regions because removal of the easily hydrolyzed material has little effect on the diffraction pattern.

Hemicellulose is a mixture of amorphous branched-chain polysaccharides consisting of a few hundred sugar residues. They are easily hydrolyzed to monomeric sugars and uronic and acetic acids. Many different hemicelluloses have been isolated from wood.

Extractives and Ash. The amount of extractives in wood varies from 5 to 20% by weight and includes a wide variety of organic chemicals (11). Many of these function as intermediates in tree metabolism as energy reserves or participate in the tree's defense mechanism against microbiological attack. The extractives contribute to wood properties such as color, odor, and decay resistance.

The ash content is 0.2–0.5% by weight for temperate woods and 0.5–2.0% by weight for tropical woods. The principal elemental components of wood ash are calcium and potassium with lesser amounts of magnesium, sodium, manganese, and iron. Carbonate, phosphate, silicate, oxalate, and sulfate are likely anions. Some woods, especially from the tropics, contain significant amounts of silica.

The chemical compositions of selected North American hardwoods and softwoods are given in Table 1 (10).

Table 1. Chemical Composition of Some North American Woods^a

Scientific name	Common name	Glucan	Xylan	Galactan	Arabinan	Mannan	Uronic anhydride	Acetyl	Lignin	Ash
<i>Hardwoods (Angiosperms)</i>										
<i>Acer rubrum</i> L.	Red maple	46	19	0.6	0.5	2.4	3.5	3.8	24	0.2
<i>Acer saccharum</i> Marsh.	Sugar maple	52	15	<0.1	0.8	2.3	4.4	2.9	23	0.3
<i>Betula alleghaniensis</i> Britton	Yellow birch	47	20	0.9	0.6	3.6	4.2	3.3	21	0.3
<i>Betula papyrifera</i> Marsh.	White birch	43	26	0.6	0.5	1.8	4.6	4.4	19	0.2
<i>Fagus grandifolia</i> Ehrh.	Beech	46	19	1.2	0.5	2.1	4.8	3.9	22	0.4
<i>Liquidambar styraciflua</i> L.	Sweetgum	39	18	0.8	0.3	3.1			24	0.2
<i>Platanus occidentalis</i> L.	American sycamore									
Fast growth		44	18	2.0	0.7	2.2	5.6	5.3	20	0.8
Slow growth		43	15	2.2	0.6	2.0	5.1	5.5	23	0.7
<i>Populus deltoides</i> Bartr. ex Marsh.	Eastern cottonwood									
Fast growth		42	19	1.3	0.5	2.9	5.5	4.0	24	0.7
Slow growth		47	15	1.4	0.6	2.9	4.8	3.1	24	0.8
<i>Populus tremuloides</i> Michx.	Quaking aspen	49	17	2.0	0.5	2.1	4.3	3.7	21	0.4
<i>Quercus falcata</i> Michx.	Southern red oak	41	19	1.2	0.4	2.0	4.5	3.3	24	0.8
<i>Ulmus americana</i> L.	White elm	52	12	0.9	0.6	2.4	3.6	3.9	24	0.3
<i>Softwoods (Gymnosperms)</i>										
<i>Abies balsamea</i> (L.) Mill.	Balsam fir	46	6.4	1.0	0.5	12	3.4	1.5	29	0.2
<i>Gingo biloba</i> L.	Ginko	40	4.9	3.5	1.6	10	4.6	1.3	33	1.1
<i>Juniperus communis</i> L.	Common juniper	41	6.9	3.0	1.0	9.1	5.4	2.2	31	0.3
<i>Larix decidua</i> Mill.	Common larch (sapwood)	46	6.3	2.0	2.5	11	4.8	1.4	26	0.2
<i>Larix laricina</i> (Du Roi) K. Koch	Tamarack	46	4.3	2.3	1.0	13	2.9	1.5	29	0.2
<i>Picea abies</i> (L.) Karst.	Norway spruce	43	7.4	2.3	1.4	9.5	5.3	1.2	29	0.5
<i>Picea glauca</i> (Moench) Voss	White spruce	45	9.1	1.2	1.5	11	3.6	1.3	27	0.3
<i>Picea mariana</i> (Mill.) B.S.P.	Black spruce	44	6.0	2.0	1.5	9.4	5.1	1.3	30	0.3
<i>Picea rubens</i> Sarg.	Red spruce	44	6.2	2.2	1.4	12	4.7	1.4	28	0.3
<i>Pinus banksiana</i> Lamb.	Jack pine	46	7.1	1.4	1.4	10	3.9	1.2	29	0.2
<i>Pinus radiata</i> D. Don	Australian radiata ^b	42	6.5	2.8	2.7	12	2.5	1.9	27	0.2
<i>Pinus resinosa</i> Ait.	Red pine	42	9.3	1.8	2.4	7.4	6.0	1.2	29	0.4
<i>Pinus rigida</i> Mill.	Pitch pine	47	6.6	1.4	1.3	9.8	4.0	1.2	28	0.4
<i>Pinus strobus</i> L.	Eastern white pine	45	6.0	1.4	2.0	11	4.0	1.2	29	0.2
<i>Pinus sylvestris</i> L.	Scots pine	44	7.6	3.1	1.6	10	5.6	1.3	27	0.4
<i>Pinus taeda</i> L.	Loblolly pine	45	6.8	2.3	1.7	11	3.8	1.1	28	0.3
<i>Pseudotsuga menziesii</i> (Mirb.) Franco	Douglas-fir	44	2.8	4.7	2.7	11	2.8	0.8	32	0.4
<i>Thuja occidentalis</i> L.	Northern white cedar	43	10.0	1.4	1.2	8.0	4.2	1.1	31	0.2
<i>Tsuga canadensis</i> (L.) Carr.	Eastern hemlock	44	5.3	1.2	0.6	11	3.3	1.7	33	0.2

^a The values expressed are for percent oven-dry wood and extractive-free wood.

^b Australian-grown wood. Percent oven-dry wood.

Wood–Liquid Relationship

Adsorption. Wood is highly hygroscopic. The amount of moisture adsorbed depends mainly on the relative humidity and temperature (Fig. 1). Exceptions occur with species with high extractive contents (eg, redwood, cedar, and teak). The equilibrium moisture contents of such woods are generally

somewhat lower than those given in Figure 1.

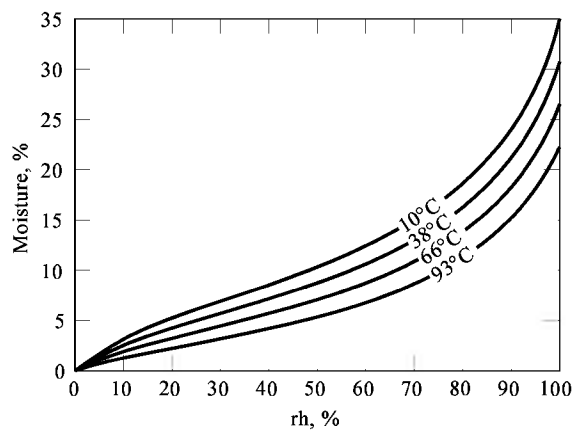


Fig. 1. Relationship between the moisture content of wood (% of dry wood) and relative humidity at different temperatures.

In green wood, the cell walls are saturated, whereas some cell cavities are completely filled and others may be completely empty. Moisture in the cell walls is called bound, hygroscopic, or adsorbed water. Moisture in the cell cavities is called free or capillary water. The distinction is made because, under ordinary conditions, the removal of the free water has little or no effect on many wood properties. On the other hand, the removal of the cell wall water has a pronounced effect.

At equilibrium with relative humidity below 100%, the moisture in wood is present primarily in the cell walls. The moisture content at which the cell walls would be saturated and the cell cavities empty is called the fiber saturation point. Actually, such distribution is impossible. Beginning at ~90% relative humidity, some condensation may occur in small capillaries. The determination of the fiber saturation point is based on the fact that certain properties of wood (eg, strength and volume) change uniformly at first with increasing moisture content and then become independent of the moisture content (Fig. 2). The equilibrium moisture content (usually determined by extrapolation), at which the property becomes constant at 25 to 30% moisture, is represented by the fiber saturation point.

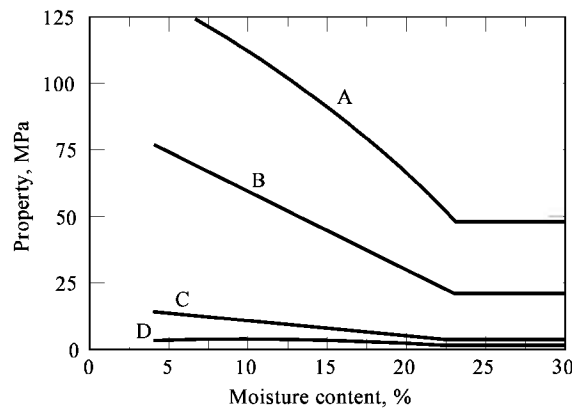


Fig. 2. Relationship between various strength properties of clear wood and moisture content. A, modulus of rupture; B, compression parallel to grain; C, compression perpendicular to grain; D, tension perpendicular to grain.

The density of wood substance is ~1.5 g cm⁻³. A species with a density of 0.5 g cm⁻³ (based on oven-dry weight and volume) has a void volume of 66.6%. Each 100 kg of the totally dry wood occupies 0.2 m³ (7 ft³) and contains 0.067 m³ (2.4 ft³) wood substance and 0.133 m³ (4.7 ft³) void. In waterlogged condition, the cell walls adsorb ~0.028 m³ (1 ft³) water, whereas the cell cavities contain ~0.133 m³ (4.7 ft³) water.

The average specific gravity of different species is given in Table 2 (6). The conventional way for expressing the specific gravity of wood is in terms of the oven-dry weight and volume at 12% moisture content. The specific gravity based on the volume of the oven-dry wood is ~6% higher.

Table 2. Specific Gravity of Some Common Woods Growing in the United States

Wood	Specific gravity ^a
aspen	0.40
birch, yellow	0.67
cottonwood, eastern	0.42
Douglas-fir, coast-type	0.51
fir, balsam	0.38
hemlock, western	0.44
maple, sugar	0.68
oak, white	0.73
pine	
lodgepole	0.43
ponderosa	0.42
longleaf	0.62
spruce, Engelmann	0.36
walnut, black	0.59

^a Based on oven-dry weight and volume at 12% moisture content.

At low relative humidities, adsorption is due to interaction of water with accessible hydroxyl groups. These are present on the lignin and on the carbohydrates in the noncrystalline or poorly crystalline regions. The high differential heat of adsorption by dry wood, ~1.09 kJ g (469 Btu lb) water,

reflects a very high affinity for moisture (12–14).
At high relative humidities, adsorption is believed to occur in response to a tendency for cellulose chains and lignin to disperse (solution tendency). Complete dispersion (dissolution) is prevented because of the strong interchain or interpolymer bonding at certain sites or regions. The differential heats of adsorption are much smaller than at low relative humidities.

Because the relative humidity of the atmosphere changes, the moisture content of wood undergoes corresponding changes. Effective protection against fluctuating atmospheric conditions is furnished by a surface coating of certain finishes, provided the coating is applied to all surfaces of wood through which moisture gains access. However, no coating is absolutely moistureproof; coatings simply retard the rate at which moisture is taken up from or given off to the atmosphere. This means that coatings cannot be relied upon to keep moisture out of wood that is exposed to dampness constantly or for prolonged periods. Coatings vary markedly in their moisture-retarding efficiencies.

For some uses, it is important to protect wood against water, eg, doors, windows, door and window frames, and the lap and butt joints in wood siding. Water repellents and water-repellent preservatives, long used in the millwork industry, provide protection from wetting. They are designed to penetrate into wood, but they leave a very thin coating of wax, resin, and oil on the surface, which repels water. However, they are not as effective in resisting water vapor (see WATERPROOFING).

Neither coatings nor water repellents alter the equilibrium moisture content or equilibrium swelling of wood. This can be accomplished only by depositing bulking agents that block normal shrinkage within the cell walls, chemically replacing the hygroscopic hydroxyl groups of cellulose and lignin with less hygroscopic groups, or forming chemical crosslinks between the structural units of wood.

Shrinking and Swelling. The adsorption and desorption of water in wood is accompanied by external volume changes. At moisture contents below the fiber saturation point, the relationship may be a simple one, merely because the adsorbed water adds its volume to that of the wood, or the desorbed water subtracts its volume from the wood. The relationship may be complicated by the development of stresses. Theoretically, above the fiber saturation point, no volume change should occur with a change in the moisture content. Actually, owing to the development of stresses, changes in volume or shape may occur. The magnitude of such stresses is minimized by drying wood under carefully controlled and empirically established conditions (13).

In the absence of drying stress (ie, with small specimens and extremely slow drying), the degree of shrinkage from the green to oven-dry condition is, as a first approximation, proportional to the specific gravity of the wood. The value of the slope of the linear relationship is equal to the average fiber saturation point of the wood. Serious deviations from the linear relationship may occur with species high in extractives.

Swelling or shrinking of wood is highly anisotropic. Tangential swelling (occurring tangent to the rings) is 1.5 to 3.5 times greater than radial swelling (occurring along a radius of the rings). Longitudinal swelling (occurring in the direction of tree growth) is usually very small. In certain abnormal woods, however, such as compression or tension wood, longitudinal swelling or shrinking may be relatively high (up to 1–2° for tension wood and 5–6° for compression wood) (14).

Permeability. Although wood is a porous material (60–70° void volume), its permeability (ie, flow of liquids under pressure) is extremely variable. This is due to the highly anisotropic shape and arrangement of the component cells and to the variable condition of the microscopic channels between cells. In the longitudinal direction, the permeability is 50 to 100 times greater than in the transverse direction (13). Sapwood is considerably more permeable than heartwood. In many instances, the permeability of the heartwood is practically zero. A rough comparison, however, may be made on the basis of heartwood permeability, as shown in Table 3.

Table 3. Relative Permeability of the Heartwood of Some Common Species, Decreasing from Group 1 to Group 4

Group 1	Group 2	Group 3	Group 4
ponderosa pine	coastal Douglas-fir	eastern hemlock	alpine fir
basswood	jack pine	Engelmann spruce	Douglas-fir ^a
red oaks	loblolly pine	lodgepole pine	tamarack
slippery elm	longleaf pine	noble fir	western redcedar
tupelo gum	western hemlock	Sitka spruce	black locust
white ash	cottonwood	western larch	red beech
	aspen	white fir	red gum
	silver maple	white spruce	white oaks
	sugar maple	rock elm	
	yellow birch	sycamore	

^a Growing in interior regions; permeability practically zero.

Transport. Wood is composed of a complex capillary network through which transport occurs by capillarity, pressure permeability, and diffusion. A detailed study of the effect of capillary structure on the three transport mechanisms is given in Stamm (13).

Drying. The living tree holds much water in its cells. A southern pine log, 5 m long and 0.5 m in diameter, for example, may weigh as much as 1000 kg and contain ~47% or 0.46 m³ (16 ft³) water.

There are a number of important reasons for drying; it reduces the likelihood of stain, mildew, or decay developing in transit, storage, or use; the shrinkage that accompanies drying can take place before the wood is put to use; wood increases in most of its strength properties as it dries below the fiber saturation point (30° moisture content); the strength of joints made with fasteners, such as nails and screws, is greater in dry wood than in wet wood dried after assembly; the electrical resistance of wood increases greatly as it dries; dry wood is a better thermal insulating material than wet wood; and the appreciable reduction in weight that accompanies drying reduces shipping costs (see DRYING).

Ideally, the temperature and relative humidity during drying should be controlled; if wood dries too rapidly, it is likely to split, check, warp, or honeycomb because of stresses. If wood dries too slowly, it is subject to development of stain and mold growth.

Air drying is a process of stacking lumber outdoors to dry (15). Control of drying rates is limited and great care must be taken to avoid degrading the wood. Drying time is a function of climatic condition; in cold temperatures or in damp coastal areas, wood dries slowly, whereas in warm temperatures and in the arid regions of the Southwest, wood dries rapidly. Typical air drying times for 2.5-cm-thick lumber of various species are shown in Table 4.

Table 4. Approximate Air-Drying and Kiln-Drying Periods for 2.5-cm Lumber

Species	Days required to	
	Air-dry to 20°	Kiln-dry to 6° ^a
baldcypress	100–300	10–20
hickory	70–200	7–15
magnolia	60–150	10–15
oak		
red	100–300	16–28
white	150–300	20–30
pine, southern	40–150	3–5
sweetgum	70–300	10–25

sycamore	70–200	6–12
tupelo	70–200	6–12
yellow poplar	60–150	6–10

^a From 20° moisture content.

Kiln drying is a controlled drying process widely used for drying both hardwoods and softwoods. Dry-bulb temperatures for hardwood lumber seldom exceed 87°C and then only at the end of the drying schedule. Dry-bulb temperatures for softwood lumber are sometimes as high as 115°C. In the initial stages, the relative humidity is maintained at a high level to control the moisture gradient in the wood and thus prevent splitting and checking. Modern kiln installations use forced-air circulation and are equipped with automatic controls for both dry- and wet-bulb temperatures. The kilns are vented to exhaust the moisture evaporated from the wood. Most kilns are steam heated, although furnace-type kilns fired with gas or oil are now common for certain softwood species. Many species, especially hardwoods, are first air dried to about 20° moisture content, then kiln dried to the moisture content at which they will be used (Table 4). A typical time schedule for kiln drying a softwood is shown in Table 5. Schedules for drying hardwoods are generally more complex (16).

Table 5. Typical Softwood Kiln-Drying Time Schedules for 2.5-cm Ponderosa Common Pine

Hours in kiln	Dry-bulb temperature, °C	Wet-bulb temperature, °C	Relative humidity, %
heartwood ^a			
1–8	54	43	52
8–16	58	44	48
16 until dry	60	44	41
sapwood ^b			
1–12	54	43	52
12–24	58	44	48
24 until dry	60	44	41

^a Kiln-dried to average 12° moisture content in 24–36 h.

^b Kiln-dried to average 12° moisture content in 48–72 h.

Special drying methods, such as superheated steam, solvent, vacuum, infrared radiation, and high frequency dielectric and microwave heating, are occasionally employed when accelerated drying is desired and the species being dried can withstand severe conditions without damage. None of these methods is of significant commercial importance.

Structural Material

Strength and Related Properties. In the framing of a building or the construction of an industrial unit, where wood is used because of its unique physical properties, strength and stiffness are primary requirements. Different species of wood have different mechanical properties that relate to the amount of wood substance per unit volume, ie, its specific gravity. Heavy wood such as oak tends to be stronger and stiffer than a light wood such as spruce. The strength of a piece of lumber depends also upon its grade or quality. The strength values of a lumber grade depend upon the size and number of such characteristics as knots, cross grain, shakes, splits, and wane (17). Wood free from these defects is known as clear wood.

Most strength properties of clear wood improve markedly as moisture is reduced below ~30%, based on the oven-dry weight (18) (Fig. 2). However, some properties (eg, tensile strength parallel and perpendicular to the grain) may be reduced if the wood is overly dried (19). In structural lumber containing defects, the improvement in mechanical properties normally associated with wood as it dries may be partially offset by degradation during drying, particularly with low grade material (Fig. 3) (20). The strength properties of wood that is conventionally kiln dried are not significantly different from those of wood that is carefully air dried. However, there is increasing interest in high temperature drying (110–120°C) of structural lumber to reduce processing time and energy consumption. These processes may reduce strength up to 20° depending upon species and property (21).

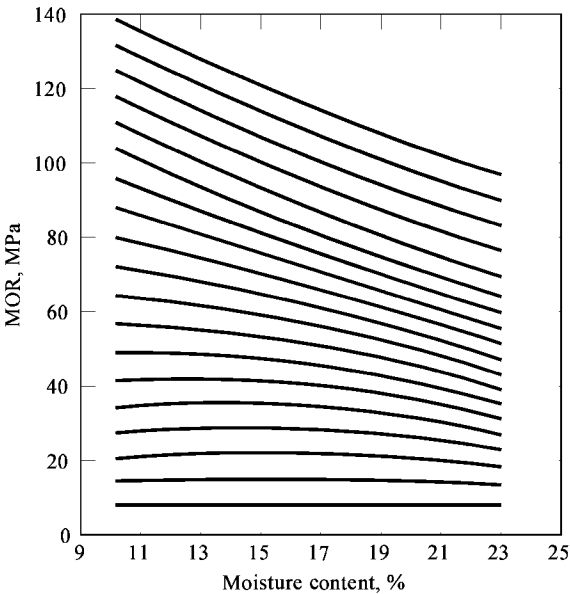


Fig. 3. Effect of moisture content on bending strength of structural lumber, MOR = modulus of rupture.

The mechanical properties of wood tend to increase when it is cooled and to decrease when it is heated (6,18). If untreated wood heated in air is not exposed to temperatures of more than ~70°C for more than about 1 year, the decrease in properties with increasing temperature is referred to as immediate or reversible; ie, the property would be lower if tested at the higher temperature but would be unchanged if heated and then tested at room temperature. The immediate effect of temperature on strength and modulus of elasticity of clear wood, based on several different loading modes, is illustrated in Figures 4–6 (6).

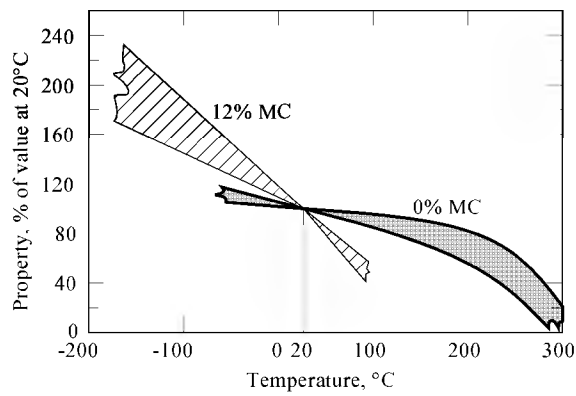


Fig. 4. The immediate effect of temperature on the modulus of elasticity of clear wood, relative to the value at 20°C. The plot is a composite of studies on the modulus as measured in bending, in tension parallel to grain, and in compression parallel to grain. Variability in reported results is illustrated by the width of the bands. MC = moisture content.

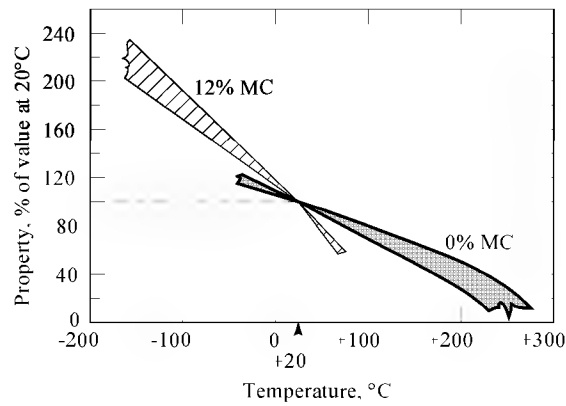


Fig. 5. The immediate effect of temperature on strength properties of clear wood, expressed as percentage of value at 20°C. Trends illustrated are composites from studies on three strength properties: modulus of rupture in bending, tensile strength perpendicular to grain, and compressive strength parallel to grain. Variability in reported results is illustrated by the width of the bands. MC = moisture content.

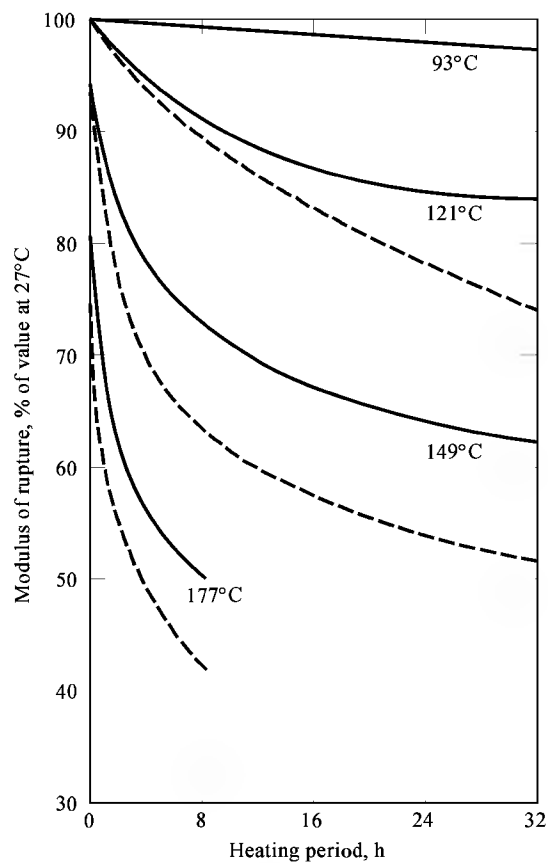


Fig. 6. Permanent effect of heating in water (—) and in steam (---) on the modulus of rupture. Data based on tests of clear Douglas-fir and Sitka spruce.

Higher temperatures result in permanent degradation. The amount of this irreversible loss in mechanical properties depends upon moisture content, heating medium, temperature, exposure period, and, to some extent, species. The effects of these factors on modulus of rupture, modulus of elasticity, and work to maximum load are illustrated in Figures 6–9 (6). The effects may be less severe for commercial lumber than for clear wood heated in air (Fig. 10). The permanent property losses shown are based on tests conducted after specimens were cooled to ~24°C and conditioned to a moisture content of

7–12% . If tested hot, presumably immediate and permanent effects would be additive. Exposure to elevated temperature for extended periods has immediate and permanent effects and must be considered in the design of structures such as chemical storage tanks.

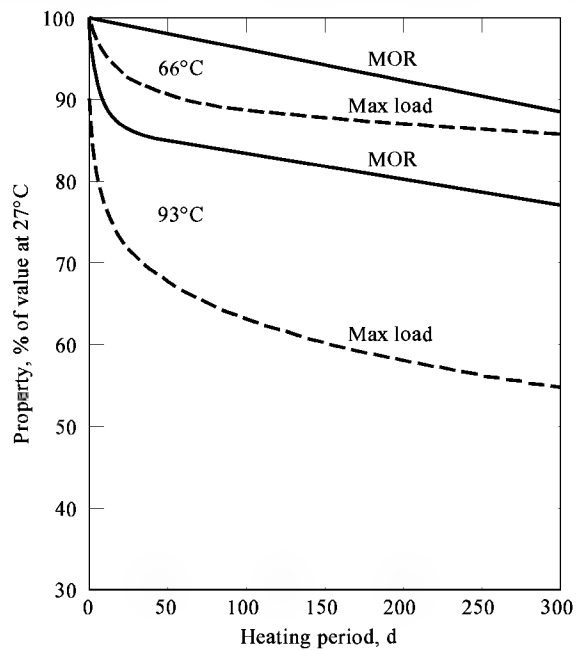


Fig. 7. Permanent effect of heating in water on work-to-maximum-load and on modulus of rupture (MOR). Data based on tests of clear Douglas-fir and Sitka spruce.

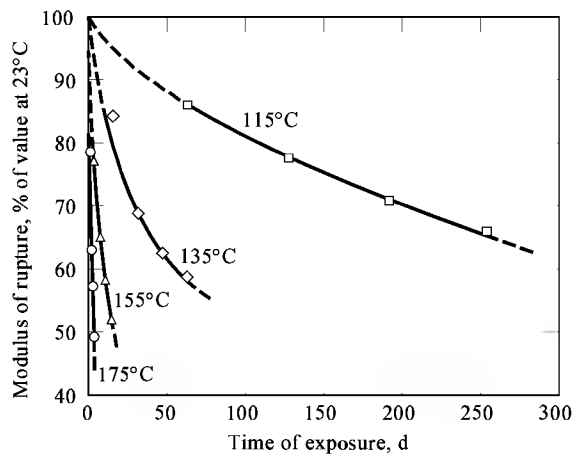


Fig. 8. Permanent effect of oven heating at four temperatures on the modulus of rupture of clear wood, based on four softwood and two hardwood species.

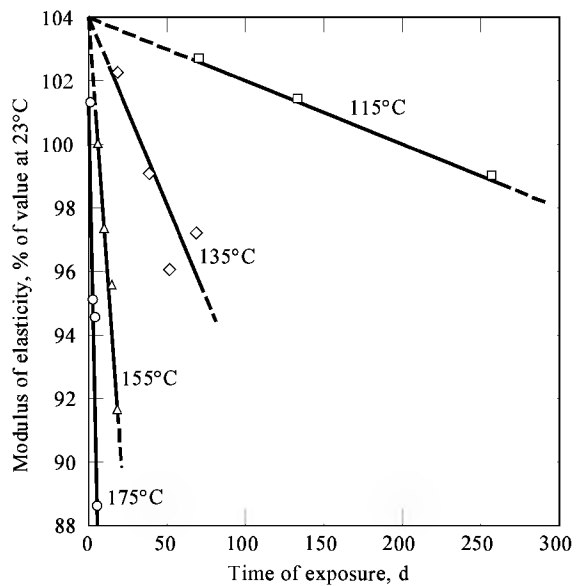


Fig. 9. Permanent effect of oven heating at four temperatures on modulus of elasticity of clear wood, based on four softwood and two hardwood species.

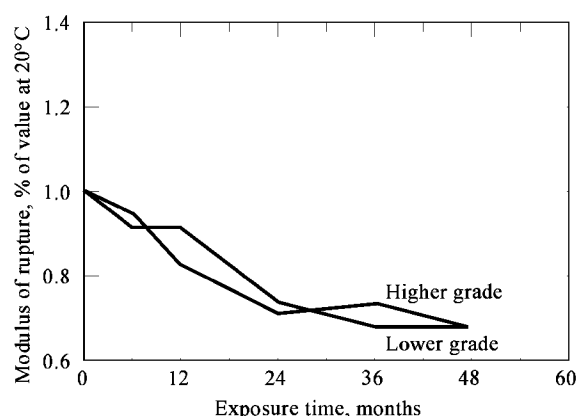


Fig. 10. Permanent effect of heating in air on bending strength of spruce-pine-fir lumber.

Repeated exposure to elevated temperature has a cumulative effect. For example, at a given set of exposure conditions, the property losses are about the same after six exposure periods of one month each and after a single 6-month period.

The shape and size of wood pieces are important in analyzing the influence of temperature. If exposure is for only a short time and the inner parts of a large piece do not reach the temperature of the surrounding medium, the immediate effect on the strength of inner parts is less than for outer parts. The type of loading must also be considered. If the piece is to be stressed in bending, its outer fibers are subjected to the greatest load and ordinarily govern the ultimate strength; under such a loading condition, the inner part having a lower temperature may be of little significance.

For extended, noncyclic exposures, it can be assumed that the entire piece reaches the temperature of the heating medium and is, therefore, subject to permanent strength losses throughout the piece, regardless of size and mode of stress application. Because dry wood is a good insulator, it often does not reach the daily extremes in temperature of the air around it in ordinary construction; thus, estimates of long-term effects should be based on the actual wood temperatures experienced by critical structural parts.

The effect of absorption of various liquids upon the strength properties of wood largely depends on the chemical nature and reactivity of the absorbed liquid. In general, neutral, nonswelling liquids have little if any effect upon the strength properties (22,23). Any liquid that causes wood to swell causes a reduction in strength. This effect may be temporary, existing only while the liquid remains in the wood. This effect may also be permanent, as in the case of chemically reactive liquids (strong acids and bases), with the magnitude being dependent upon the time and temperature of exposure and concentration of the solution.

Wood preservatives are applied either from an oil system, such as creosote, petroleum solutions of pentachlorophenol, or copper naphthanate, or a water system. Oil treatments are relatively inert with wood material, and thus, have little effect on mechanical properties. However, most oil treatments require simultaneous thermal treatments, which are specifically limited in treating standards to preclude strength losses (24).

The mechanical properties of wood can be damaged by preservative systems. In North America, the primary waterborne preservatives are copper-based systems, which may contain chromic acid and arsenic, ammoniacal or amine-based copper systems with supplemental zinc, arsenate, or quaternary ammonium chloride. The effects of waterborne preservative treatments are directly related to key pretreatment, treatment, and post-treatment processing factors (25). Because treating standards have strict thermal limits to avoid affecting strength, few design modifications are required to allowable design stresses if waterborne preservative-treated materials are treated to standards (24). The two design exceptions involve (a) restrictions on impact-load adjustments for wood treated with waterborne preservatives and (b) adjustment for the mechanical process of making knife-like incisions, which is used with difficult-to-treat species to improve preservative penetration and distribution. Incising reduces strength; the literature supports a 5–10% reduction in MOE and a 20–30% reduction in allowable design stresses (F_b , F_t , F_c , F_v).

Most commercial fire retardant treatment formulations are proprietary. Fire retardant treatments generally reduce allowable design stresses by 10–25%. The magnitude of these reductions varies depending upon the fire retardant chemical, the severity of treatment and processing conditions, and the property being considered (25). Treatment standards specifically limit thermal processing to control design adjustments to the levels stated above (24). Specific modification factors to allowable design stresses depend on the specific fire retardant treatment. The treater should be consulted to obtain appropriate modification factors for specific fire retardant treatments.

Another problem occurs when some fire retardant formulations are exposed to elevated temperatures (eg, when used as roof trusses or as roof sheathing); thermal-induced strength reductions can occur in-service. The thermo-chemical factors were discussed by LeVan and Winandy (26), and a kinetic degrade model was developed (27). The treater should be consulted to obtain appropriate in-service modifications for specific fire retardant treatments.

Reaction to Heat and Fire. The physical and chemical properties of wood, like those of any organic material, are subject to deterioration. The rate and extent of deterioration are governed by the interdependent factors of temperature, time, and moisture. In locations not conducive to decay or insect attack, wood is extremely stable at ordinary temperatures. However, with increasing temperature, the degradation of surface layers progresses into the interior layers. Prolonged heating at temperatures as low as 90°C may cause charring.

In general, the thermal degradation of wood and other cellulosic substances proceeds along one of two competing reaction pathways (28). At temperatures up to ~200°C, carbon dioxide and traces of organic compounds are formed, in addition to the release of water vapor. The gases are not readily ignitable, but under certain conditions, a pilot flame can ignite the volatiles after 14 to 30 min at 180°C (29). Exothermic reactions may occur near 200°C and, in situations where heat is conserved, self-ignition at temperatures as low as 100°C has been observed (30). Times and temperatures that might result in smoldering initiation can be determined (31). To provide a margin of safety, 77°C should be the upper limit in prolonged exposure near heating devices.

Temperatures in excess of 200°C lead to much more rapid decomposition. Under these conditions, the pyrolysis gases contain 200 or more different components (32–34) and the degradation is accompanied by reduction in weight, depending on temperature and duration of heating (35) (Fig. 11). Thermogravimetric analysis of wood [α -cellulose and lignin (Fig. 12)] indicates that a slow initial weight loss for lignin and wood begins at ~200°C.

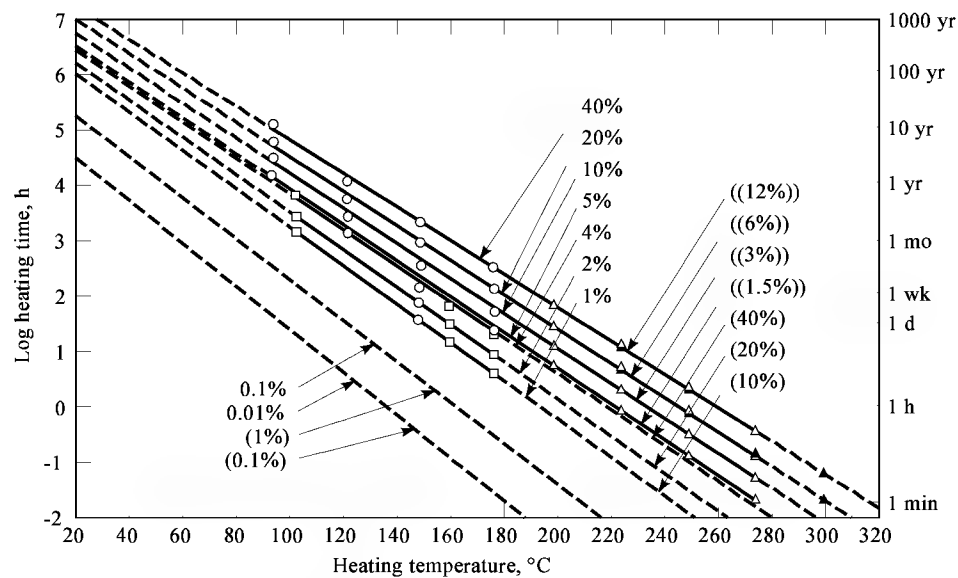


Fig. 11. Logarithm of heating time compared with temperature to attain various degrees of degradation of wood (35). No parentheses indicates weight loss on oven heating; a single parenthesis indicates modulus of rupture on oven heating; and double parentheses indicates weight loss on heating beneath surface of molten metal.

Courtesy of Industrial and Engineering Chemistry.

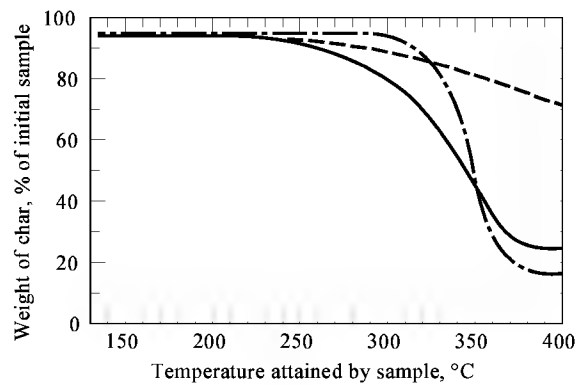


Fig. 12. Portion of dynamic thermogravimetric curves between 130 and 400 °C, with temperature rising 6 °C per minute for wood, lignin, and α-cellulose. — wood, --- lignin, ·-·-· α-cellulose (36).

Differential thermal analyses (dta) of wood and its components indicate that the thermal degradation reactions in an inert atmosphere release less than 5% of the heat released during combustion in air. The typical dta data given in Figure 13 (36) show two main exothermic peaks for wood. The peak near 320 °C represents the flaming reaction resulting from combustion of the volatiles associated with cellulose pyrolysis; the other peak, near 440 °C, represents glowing in place of the solid charcoal residue. The dta for α-cellulose and lignin in Figure 11 can be superimposed to approximate the curve for wood. Several extensive reviews on thermal degradation of wood are available (37–40).

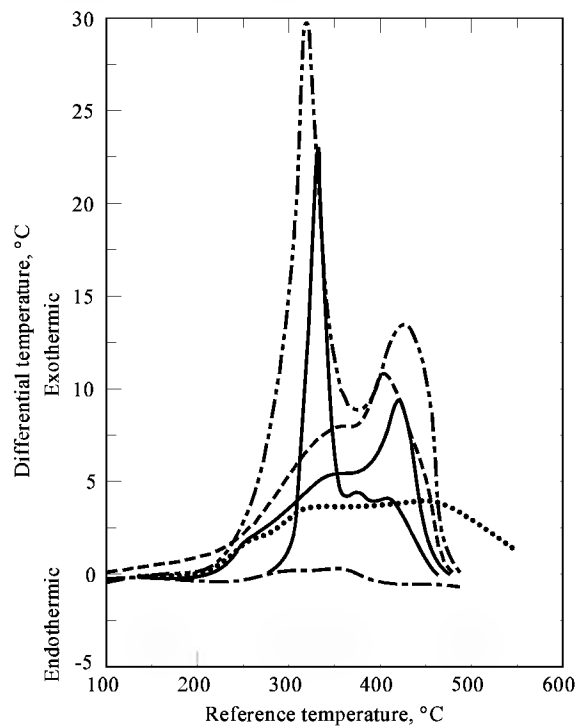


Fig. 13. Differential thermal analysis of wood and its components at a heating rate of 12 °C per minute and a gas flow rate of 30 cm³ per minute. Sample

weight: wood 40 mg, α -cellulose 20 mg, lignin 10 mg, charcoal 12 mg. Treated wood sample contains 9% by weight of commercial fire retardant. — untreated wood in O₂; --- untreated wood charcoal in O₂; -- untreated wood in N₂; ——— untreated α -cellulose in O₂; untreated sulfate lignin in O₂; ---- wood treated with fire retardant in O₂.

Building code requirements for fire performance are mainly concerned with noncombustibility (41), fire endurance (42,43), and surface burning characteristics (44). Wood, even in its treated form, does not meet the building code requirements for a noncombustible material. However, for some specific applications where noncombustible materials are required, the codes permit the substitution of fire retardant treated wood.

Wood in its untreated form has good resistance or endurance to fire penetration when used in thick sections for walls, doors, floors, ceilings, beams, and roofs. This endurance is due to low thermal conductivity, which reduces the rate at which heat is transmitted to the interior. Typically, when the fire temperature at the surface of softwood is 870–980°C, the inner char zone temperature is ~290°C, and 6 mm further inward, the temperature is 180°C or less. The penetration rate of this char line is ~0.6 mm min, depending on the species, moisture content, and density (45,46). Owing to this slow penetration rate and low thermal conductivity, large wood members retain a substantial portion of their load carrying capacity for considerable time during fire exposure (47).

The surface burning characteristics (flame spread index and smoke developed index) for wood and wood products as measured by American Society for Testing and Materials (44) can be reduced with fire retardant treatments, either chemical impregnation or coatings (48). Fire retardant treatments also reduce the heat release rate of a burning piece of wood (49,50). The heat release rates (51) of the burning materials are an important factor in fire growth.

Fire retardant chemicals, such as ammonium phosphate, ammonium sulfate, zinc chloride, guanylurea phosphate, dicyandiamide phosphate, borax, and boric acid, are often used in combinations. Borax and boric acid mixtures are moderately effective in reducing flamespread and afterglow without premature charring during severe drying operations. Although very hygroscopic, zinc chloride is an effective flame retardant; boric acid is often added to retard afterglow. Fire retardant treatments can adversely affect the strength properties of wood. Elevated temperatures in service can cause further strength loss (52). Fire retardants such as ammonium sulfate can have a corrosive effect on metal fasteners. In exterior applications, a treatment with resistance to weathering and leaching is important (53,54).

Solutions of these fire retardant formulations are impregnated into wood under full cell pressure treatment to obtain dry chemical retentions of 65 to 95 kg m⁻³; this type of treatment greatly reduces flame-spread and afterglow. These effects are the result of changed thermal decomposition reactions that favor production of carbon dioxide and water (vapor) as opposed to more flammable components (55). Char oxidation (glowing or smoldering) is also inhibited.

Some of the chemicals mentioned above and others, such as chlorinated rubber or paraffin, antimony trioxide, calcium carbonate, calcium borate, pentaerythritol, alumina trihydrate, titanium dioxide, and urea–melamine–formaldehyde resin, may be used to formulate fire retardant coatings. Many of these coatings are formulated in such a way that the films intumesce (expand) when exposed to fire, thus insulating the wood surface from further thermal exposure. Fire retardant coatings are mostly used for existing construction.

Resistance to Chemicals. Different species of wood vary in their resistance to chemical attack. The significant properties are believed to be inherent to the wood structure, which governs the rate of ingress of the chemical and the composition of the cell wall, which affects the rate of action at the point of contact (56).

Wood is widely used as a structural material in the chemical industry because it is resistant to a large variety of chemicals. Its resistance to mild acids is far superior to that of steel but not as good as some of the more expensive acid-resistant alloys. Wood tanks used to store cold, dilute acid have a relatively long service life. However, increasing concentration or temperature causes the wood tank to deteriorate rapidly (6).

Softwoods are generally more resistant to acids than are hardwoods because they have high lignin and low hemicellulose contents. In general, heartwood is more resistant to acids than sapwood, probably because of heartwood's higher extractive content and slower movement of liquid into the heartwood. For these reasons, the heartwood of certain conifers has been widely used in the chemical industry.

Oxidizing acids, such as nitric acid, attack wood faster than common mineral acids, although wood is frequently used in contact with dilute nitric acid. Oxidizing acids not only attack wood by hydrolysis of the polysaccharides but also degrade these polymers through oxidative reaction. Wood shows excellent resistance to organic acids, which gives it a distinct advantage compared with steel, concrete, rubber, and some plastics. Mild organic acids such as acetic acid have little effect on wood strength.

Alkaline solutions attack wood more rapidly than acids of equivalent concentrations, whereas strong oxidizing chemicals are harmful. Wood is seldom used where resistance to chlorine and hypochlorite solutions is required. These chemicals cause extensive degradation of cell wall polymers. Wood tanks are, however, satisfactory for holding hydrogen peroxide solutions and give good service on contact with strong brine. Solutions of iron salts cause degradation, particularly of the polysaccharides.

In contact with iron under damp conditions, wood may show severe deterioration within a few years (57). Species high in acidic extractives seem especially prone to such attack.

Because traces of iron reduce the brilliance of many dyes, wood tanks have long been preferred to steel in the manufacture of dyes. Similarly, vinegar and sour foodstuffs are processed in wood tanks because common metals impart a metallic taste. Ease of fabrication may be the reason for using wood tanks in less accessible areas to which ready-made tanks of other materials cannot be easily moved.

Resistance to chemical attack is generally improved by resin impregnation, which protects the undying wood and reduces movement of liquid into the wood. Resistance to acids can be obtained by impregnating with phenolic resin and to alkalis by impregnating with furfural resin (see PHENOLIC RESINS).

Biodeterioration. The principal organisms that degrade wood are fungi, bacteria, insects, and marine borers. Decay, molds, and stain are caused by fungi. Decay is the most serious kind of damage because it causes structural failure and consequently, tremendous economic losses. Soft rot is another type of decay that weakens wood, but it typically progresses slowly and is most often associated with very wet wood. Moisture conditions conducive to decay occur when the moisture content of the wood is above fiber saturation (~30%). The optimum temperature range for most decay fungi is about 25–30°C, although some species grow at temperatures as low as 0°C and some as high as 45°C. The optimum pH is in the range of 4.5 to 5.5. Oxygen is essential for growth of all species. Decay can be prevented by keeping wood either too dry (below 20% moisture content) or too wet (lumens filled with water) for fungal development, by using naturally decay-resistant species, or by treating with preservatives.

Mold and stain fungi primarily attack the sapwood. Mold fungi growth occurs primarily on the surface of the wood, while stain fungi may cause a stain throughout the affected sapwood. These fungi can be controlled by dipping the lumber in a fungicidal solution immediately after cutting.

Bacterial degradation of wood generally is not a serious problem, although in some situations of extreme wetness, bacteria may increase the permeability of wood after many years or reduce the strength of the wood (58).

Termites are the most destructive insects that attack wood. Their attack can be prevented or lessened by using naturally resistant wood or by treating wood with preservatives. For subterranean termites, which generally require contact with the ground to survive, poisoning the soil around the wood structure is the principal means of preventing infestation. A promising new approach to subterranean termite control is the use of food bait with an insecticide (59). The drywood termite flies directly to the wood, bores into it, and does not require contact with the ground. Physical barriers, such as paint or screens, prevent infestation. Despite great differences between fungi and termites, chemicals that inhibit fungi usually also inhibit termites.

Marine borers inhabit saline or brackish waters where they cause serious destruction to untreated wood. The mollusks include the Teredo and Bankia borers; among the crustaceans, the Limnoria borers are the most widespread and destructive. Preservatives or borer-resistant woods deter marine borers (see COATINGS, MARINE).

For practical purposes, the sapwood of all species may be considered to be susceptible to biodeterioration. The heartwood of some species, however, contains toxic extractives that protect it against biological attack. Among the native species that have decay-resistant or highly decay-resistant heartwood are bald cypress, redwood, cedars, white oak, black locust, and black walnut (60). Douglas-fir, several of the pines, the larches, and honey locust

are of intermediate decay resistance. Species low in decay resistance include the remainder of the pines, the spruces, true firs, ashes, aspens, birches, maples, hickories, red and black oaks, tupelo, and yellow poplar. Native woods considered somewhat resistant to termite attack include close-grained redwood heartwood and resinous heartwood of southern pine (66). Although several tropical woods show resistance to marine borers, no commercial native woods are sufficiently borer resistant to be used untreated (6).

The best protection for wood against the attack of decay fungi, insects, or marine borers is obtained by applying preservatives under pressure before installation (61,62). Both oil-type preservatives, such as creosote or petroleum solutions of pentachlorophenol, and waterborne preservatives, such as copper-chrome arsenate and ammoniacal-copper arsenate, are used when wood is to be in direct contact with the ground or in the marine environments.

Where wood is to be used under low to moderate decay hazard conditions (eg, above ground), it can be protected by brushing, spraying, dipping, or steeping (62,63). Once decay is established, preservatives brushed onto the wood will not penetrate, and decay cannot be eradicated in this way. However, high vapor pressure fungicides (fumigants) penetrate deeply into wood and have successfully stopped internal decay in structural timbers. Diffusible preservatives such as boron and fluoride are also used to eradicate decay (64).

Modified Wood

In addition to preservation or fire protection, wood is modified to reduce the rate that moisture is sorbed by the wood (water repellency) and or to reduce the shrinking and swelling that occur at equilibrium (dimensional stability) under conditions of fluctuating relative humidity. Certain species with high extractives content, especially in the cell walls, have greater water repellency and, in some cases, greater dimensional stability than species with low extractives content. This suggests a means of obtaining still greater reductions in the rate and extent of swelling and shrinking, that is, by filling the voids in the wood to reduce rate or deliberately adding large amounts of bulking agents to the cell walls to improve dimensional stability.

Low molecular weight, nonswelling vinyl-type monomers can be impregnated into woods that polymerize *in situ* into the void structure by radiation or heat and a catalyst (65). The hygroscopic characteristics of the wood substance are not altered because little, if any, polymer penetrates the cell walls. However, because of the high polymer content (70–100%), the normally high void volume of wood is greatly reduced. With the elimination of this important pathway for vapor or liquid water diffusion, the response of the wood substance to changes in relative humidity or water is very slow and moisture resistance or water repellent effectiveness (WRE) is greatly improved (66). Water repellent effectiveness is measured as follows:

Swelling or moisture uptake of control

$$\text{WRE} = \frac{\text{specimen during exposure to water for } t \text{ minutes}}{\text{Swelling or moisture uptake of treated specimen during exposure to water also for } t \text{ minutes}} \times 100$$

Hardness is increased appreciably. Wood-polymer composites are currently used in certain sporting equipment, musical instruments, decorative objects, and high performance flooring.

To improve dimensional stability, low molecular weight chemicals are used that penetrate the cell walls and either bond to the cell wall polymers or polymerize in the cell wall. Improvements in dimensional stability are measured by antishrink efficiency (ASE):

ASE

$$\left[1 - \left(\frac{\text{Percentage swelling of treated specimen}}{\text{Percentage swelling of control}} \right) \right] \times 100$$

ASE is a measure of the extent to which the swelling and shrinking tendency has been reduced at an equilibrium condition.

The wood cell wall can be bulked with leachable polyethylene glycol (PEG) to achieve an ASE of about 80% (67). In this case, the wood is usually treated in a green condition and the PEG is exchanged for the cell wall water. The green wood is soaked in a 30% by weight PEG-1000 solution for a length of time depending upon the thickness of the wood; two coats of polyurethane varnish are usually applied later to seal in the PEG and exclude water. Maximum ASE of 80% is achieved at PEG loadings of 45% weight gain. The strength properties of PEG-treated wood approximate those of untreated green wood (see GLYCOLS; POLYETHERS).

It is possible to react an organic moiety to the hydroxyl groups on cell wall components. This type of treatment also bulks the cell with a permanently bonded chemical (68). Many compounds modify wood chemically. The best results are obtained by the hydroxyl groups of wood reacting under neutral or mildly alkaline conditions below 120°C. The chemical system used should be simple and must be capable of swelling the wood structure to facilitate penetration. The complete molecule must react quickly with wood components to yield stable chemical bonds while the treated wood retains the desirable properties of untreated wood. Anhydrides, epoxides, and isocyanates have ASE values of 60–75% at chemical weight gains of 20–30%.

Thermosetting phenolic resins have been used successfully to penetrate and polymerize in the cell wall. The rate-determining step for successful treatment is the penetration of the resin into the cell wall. With green wood, this rate depends on the diffusion rate of the resin in the lumen-trapped water. With dry wood, pressure can be applied. In this case, the rate depends on the permeability of the wood. For both processes, the size of the object is very important; long treating times are used for large pieces and for heartwood. The sapwood of some species is sufficiently permeable to admit resin fairly uniformly and in a reasonable time. Thin veneers are generally pressure impregnated (up to 1.4 MPa ~ 14 atm) with a 30% aqueous solution of a water-soluble resin. The wood is then slowly dried and heated at ~ 150°C for 20 min to set the resin. Laminates are built up by gluing the individual sheets together. The product is called impreg. Its density is ~20% higher than that of the original wood, and its color is that of the original wood or slightly darker. The ASE of impreg increases with increasing content of phenolic resin and then tends to level off at ~65% when the resin content reaches 30–35%, based on original wood. Impreg generally contains 25–35% resin (67).

The mechanical properties of resin-impregnated wood are improved or not affected except for toughness, which is reduced by as much as 60% (69). Treatment with phenol–formaldehyde resins increases the decay resistance. Impreg stakes containing ~30% resin had an average service life of 12 years in ground tests (70). Biological resistance may be due to the fact that the cell walls of the treated wood resist moisture-supporting decay. It could also be due in part to toxic effects of partially polymerized phenolic resin on the destructive organism.

Heat resistance is improved markedly by resin impregnation. A block of impreg, subjected to forty-five 1-h exposures at 204°C, showed no apparent loss in properties, although an untreated sample showed signs of deterioration after three 1-h exposures (71).

The largest industrial application of impreg is in die models for automobile body parts and other model dies. The dimensional stability and ease of shaping are the reasons that impreg, although expensive, has wide application.

If pressure is applied to dry, resin-treated veneers while they are being heat cured, a densified product (1.35 kg mL) is obtained. This material, called compreg, retains most of the advantages of impreg. In addition, owing to the two- to threefold increase in density, the mechanical properties are appreciably better than those of the original wood. The strength of compreg is increased in proportion to the compression. Compreg is less tough than untreated wood but more tough than impreg.

Because of the plasticizing action of the resin-forming materials, the wood can be compressed under considerably lower pressures than dry, untreated wood. For example, treated spruce, cottonwood, and aspen veneer, dried to a moisture content of ~ 65 but not cured, are compressed, when subjected to a pressure of only 1.72 MPa (~ 17 atm) at 149°C, to about half the original thickness and a specific gravity of ~ 1.0.

In a 24-h water soaking test, compreg has an ASE value of 95%. The rate of water pickup is so slow that complete swelling equilibrium of a 1.27-cm specimen is not achieved in a year at room temperature. Compreg is brown and acquires a high polish on buffing. It is made commercially in small quantities and is used for knife handles, gears, and certain musical instruments and decorative objects.

Bending is another treatment process. Above 80°C, green wood becomes readily deformable. The new shape persists after cooling to room temperature and drying under restraint. This is the basis for commercial bending of wood to various shapes. The deformation, however, is not a fully plastic one. An elastic component persists and produces some strain recovery at high or cyclic humidification. Ammonia is also used to bend wood (72). The wooden object is immersed in liquid ammonia for a period of time, depending on the dimensions, imparting appreciable plasticity to the wood. The

ammonia is allowed to evaporate from the deformed wood, and very little strain recovery occurs on humidification.

Chemical Raw Material

Wood is one of our most important renewable biomass resources. Unlike most biomass sources, wood is available year round and is more stable on storage than other agricultural residues. In the United States, wood residues from industrial by-products totaled 60.8×10^6 metric tons in 1993 (73). Increasingly, residues are incorporated into manufactured wood products and are used as a fuel, replacing petroleum, especially at wood-industry plants (73); some is converted to charcoal but most is used in the pulp and paper industry. Residues are also available for manufacturing chemicals, generally at a cost equivalent to their fuel value (see FUELS FROM BIOMASS; FUELS FROM WASTE).

Wood can be pyrolyzed (heated to 400°C or higher in the absence of oxygen) to produce a variety of chemical compounds. For example, various wood species heated to 400°C yielded 31–41% charcoal, 3–7% acetic acid, 1.5–2.5% methanol, 11–19% tar, and 15–17% gases (74). The gases occurring are predominately hydrogen, carbon monoxide, carbon dioxide, and methane. Wood gasification takes place at ~1,000°C in the presence of a controlled amount of oxidizing agent. The product gas composition depends upon the starting moisture content of the wood. In addition to the above-mentioned gases, some low molecular weight aliphatic hydrocarbons are also produced.

Wood is about 65–75% carbohydrate and has been considered as a potential source of ethanol for fuel. The carbohydrate material can be hydrolyzed to monomer sugars, which in turn can be fermented to produce ethanol. However, wood carbohydrates are expensive to hydrolyze. Hydrolysis with acids and enzymes is impeded by the crystalline structure of cellulose. Lignin interferes with processing, and hydrolytic by-products such as furfural, acetic acid, and derivatives of lignin and extractives can inhibit fermentation. Research is still being conducted on wood hydrolysis to develop a process that is economically sound. Furfural is a useful chemical feedstock and results from the dehydration of pentose sugars. It can be obtained in 9 to 10% yield from the dilute acid hydrolysis of hardwoods (75).

The principal chemical industry based on wood is pulp and paper. In 1995, ~114.5 $\times 10^6$ metric tons of wood were converted into ~60 $\times 10^6$ metric tons of fiber products ranging from newsprint to pure cellulose in the United States (1,76). Pure cellulose is the raw material for a number of products, eg, rayon, cellulose acetate film base, cellulose nitrate explosives, cellophane, celluloid, carboxymethylcellulose, and chemically modified cellulosic material.

Most of the more than 54.5 $\times 10^6$ metric tons of organic material removed from wood during pulping are burned for their energy content and to recover the inorganic pulping chemicals. Some organic chemicals are recovered from this waste stream, referred to as black liquor. A large proportion of the organic chemicals recovered from sulfite pulp mills are lignosulfonates. Lignosulfonates are used as oil-well drilling fluids, binders for animal food pellets, and as an additive to improve the structural properties of concrete. Some of the recovered lignin-derived material is used to produce vanillin, and the recovered sugars are used to produce small amounts of yeast, ethanol, and acetic acid. The most valuable chemical by-products are isolated at kraft pulp mills. These chemicals are sulfate turpentine and tall oil (a mixture of fatty acids and rosin); dimethyl sulfide and dimethyl sulfoxide are also obtained from sulfate pulping (kraft) liquors.

There are a few minor wood-based chemical industries. After chestnut blight wiped out the American chestnut, U.S. tannin production essentially ceased. The main natural tannins, wattle and quebracho, are now imported. High U.S. labor costs and the advent of synthetic tannins make re-establishment of a U.S. tannin industry unlikely. Tannins are used in oil-well drilling muds. Tree exudates are a continuing wood-based chemical industry. Tree exudates include rubber, true carbohydrate gums (eg, acacia gum), kinos (eg, the phenolic exudates from eucalyptus), balsams (eg, Storax from *Liquidambar* spp.), and many different types of oleoresins (mixtures of a solid resin and a liquid essential oil). The most important oleoresin still collected in the United States is pine gum (rosin plus turpentine).

Wood is the raw material of the naval stores industry (77). Naval stores, so named because of their importance to the wooden ships of past centuries, consist of rosin (diterpene resin acids), turpentine (monoterpene hydrocarbons), and associated chemicals derived from pine (see TERPENOIDS). These were obtained by wounding the tree to yield pine gum, but the high labor costs have substantially reduced this production in the United States. Another source of rosin and turpentine is through extraction of old pine stumps, but this is a nonrenewable resource and this industry is in decline. The most important source of naval stores is spent sulfate pulping liquors from kraft pulping of pine. In 1995, U.S. production of rosin from all sources was estimated at under 300,000 metric tons and of turpentine at 70,000 metric tons. Distillation of tall oil provides, in addition to rosin, nearly 128,000 metric tons of tall oil fatty acids annually (78).

Hydrolysis

In the acid hydrolysis process (79–81), wood is treated with concentrated or dilute acid solution to produce a lignin-rich residue and a liquor containing sugars, organic acids, furfural, and other chemicals. The process is adaptable to all species and all forms of wood waste. The liquor can be concentrated to a molasses for animal feed (82), used as a substrate for fermentation to ethanol or yeast (82), or dehydrated to furfural and levulinic acid (83–86). Attempts have been made to obtain marketable products from the lignin residue (87) rather than using it as a fuel, but currently only carbohydrate-derived products appear practical.

When concentrated acids are used, the carbohydrates are recovered in high yields, but the problem of economically recovering the large quantities of acid used has not been solved. At the present state of development, the dilute acid processes, especially percolating and two-stage, appear more promising.

A number of commercial plants using the Scholler percolation process (88) were built in Germany, Switzerland, and the former USSR; only the latter are still in operation. In general, these plants were built to produce sugars for fermentation to ethanol and yeast (qv). Except under special circumstances, however, such a process has proved to be uneconomical. In the Scholler process, a hot dilute solution of sulfuric acid is percolated through wood chips; the solubilized sugars are carried with the solution withdrawn from the bottom of the digester. The acid concentration and temperature of the charged solution are continuously increased; the more labile hemicelluloses are hydrolyzed and removed in the early part of the cycle, whereas the resistant cellulose, yielding glucose, is hydrolyzed at the end.

A percolation process requiring less time was developed at the USDA Forest Service, Forest Products Laboratory, in Madison, Wisconsin, during World War II (89–91). For Douglas-fir, this process gives a sugar yield of 40–45%; the sulfuric acid requirement is 5% based on the dry wood weight. The concentration of the resulting sugar solution is 4–5%. Yields per 100 kg of dry wood are 20–25 L of 100% ethanol.

In two-stage processes, the hemicellulose sugars are hydrolyzed in the first stage; the solubilized material is washed from the residue, which is reimpregnated with acid and passed to the second stage where the cellulose is hydrolyzed. This process is much less energy and capital intensive than the percolation process and gives a better fractionation of the hemicellulose sugars and glucose. Enzymatic saccharification of the cellulose requires two-stage processing. Sugar yields are about the same as those obtained by percolation, but concentrations of 10–12% are reached with significantly lower acid, energy, and equipment capacity requirements. The two-stage process is not operated commercially, but during World War II, some pilot studies were done in Sweden (92). More recently, the Tennessee Valley Authority (TVA) did extensive pilot plant testing on the process with a 1-ton-per-day pilot plant (93). Much of the recent work on lignocellulose use has been directed toward the development of the two-stage process with either chemical or enzymatic hydrolysis of the cellulose (94–100).

High yields of pure glucose can be obtained by enzymatic cellulose hydrolysis. However, the enzyme is expensive and does not attack lignin-encrusted cellulose. Several processes have been studied using various organisms singly and in combination after subjecting the lignocellulose to different pretreatments (94,100–104). In the Gulf process, cellulose hydrolysis and glucose fermentation take place simultaneously in one vessel.

The Iotech steam explosion process has been used to prepare cellulosic substrates for enzymatic digestion (105,106). Wood, or other lignocellulose material, is subjected to a short prehydrolysis at high temperature and pressure and then is rapidly decompressed. The combined chemical and mechanical treatment solubilizes the hemicellulose component in hot water and the lignin in alcohols. Freed of lignin encrustation, the cellulose is highly accessible to hydrolytic enzymes. The recovered lignin is thermoplastic and may be marketable as a resin component.

Although the hydrolysis of wood to produce simple sugars has not proved to be economically feasible, by-product sugars from sulfite pulping are used to produce ethanol and to feed yeast (107). Furthermore, a hemicellulose molasses, obtained as a by-product in hardboard manufacture, can be used in cattle feeds instead of blackstrap molasses (108). Furfural can be produced from a variety of wood processing byproducts, such as spent sulfite liquor, liquors from the prehydrolysis of wood for kraft pulping, hardboard plants, and hardwood wastes (109).

Fuel Properties

The fuel properties of wood can be summarized by ultimate and proximate analyses and determination of heating value. The analytical procedures are the same as those for coal, but with some modifications. Analytical results generally vary about as much within a species as they do between species, except that softwood species generally have a higher carbon content and higher heating values than hardwood species because of the presence of more lignin and resinous materials in softwood species (see FUELS FROM WASTE).

The higher heating value of wood and bark of softwood species is usually ~21 kJ g (~9000 Btu lb) and slightly less for hardwood. The higher heating value includes the heating value of the condensed steam given off. These values are within ±5% of nearly all the values reported in the literature for specific samples. A systematic study of the heating value of a species has not been done because it requires measuring the heating value at various positions in a tree and from trees selected from the entire geographic range of the species.

Wood ash generally contains calcium, potassium, phosphorus, magnesium, and silica. Ashes recovered from burned wood are ~25% water soluble and the extract is strongly alkaline. The ash fusion temperature is in the range of 1300 to 1500°C.

The moisture content of freshly cut wood varies between species and portions of the tree. Between species, it can be 30–70° on a total weight basis (65); commonly, it is 45–50° . Within a tree, the heartwood generally has lower moisture content than the sapwood. For hardwood species, this difference is usually small; for softwood species such as Douglas-fir, the difference can be as great as 30° for heartwood compared with 50° for sapwood.

Charcoal Production

Charcoal is produced by heating wood under limited access of oxygen. When wood is heated slowly to ~280°C, an exothermic reaction occurs. In the usual carbonization procedure, heating is prolonged to 400 to 500°C in the absence of air. The term charcoal also includes charcoal made from bark.

Charcoal is produced commercially from primary wood-processing residues and low quality roundwood in either kilns or continuous furnaces. A kiln is used if the raw material is in the form of roundwood, sawmill slabs, or edgings. In the United States, most kilns are constructed of poured concrete with a capacity of 40 to 100 cords of wood and operating on a 7- to 12-d cycle. Sawdust, shavings, or milled wood and bark are converted to charcoal in a continuous multiple-hearth furnace commonly referred to as a Herreshoff furnace. The capacity is usually at least 1 ton of charcoal per hour. The yield is ~25% by weight on a dry basis.

The proximate analysis of charcoal is ~20–25% volatile matter, 70–75° fixed carbon, and 5° ash. Charcoal briquets have lower heating values than charcoal lumps, because of additives in the briquets. The higher heating value of lump charcoal is ~28 kJ kg (12,000 Btu lb). The higher heating value of briquets is 23 to 25 kJ kg (~9,900–10,800 Btu lb).

To alleviate the air pollution problem associated with charcoal kilns and furnaces, the gases from the kiln and furnaces are burned (see AIR POLLUTION CONTROL METHODS). They can be burned with additional fossil fuel to recover heat and steam (110,111), or in afterburners to nearly eliminate visible air pollution and odors (112).

Charcoal was an important industrial raw material in the United States for iron ore reduction until it was replaced by coal in the early 1880s. Charcoal production increased, however, because of the demand for the by-products acetic acid, methanol, and acetone. In 1920, nearly 100 by-product recovery plants were in operation in the United States, but the last plant ceased operation in 1969.

Charcoal production has increased since the 1940s, which reflects the use of charcoal briquets for home and recreational cooking (Table 6). The charcoal currently produced is nearly all consumed as briquets for cooking. Some charcoal is used in certain metallurgical and filtration processes and horticultural uses. In Brazil, charcoal is produced in beehive-type kilns from natural and plantation-grown trees for use as a reducing agent for iron ore because Brazil does not have abundant supplies of coking coal. In many developing countries, charcoal is preferred for domestic cooking. It is made in pit-type kilns or portable sheet metal kilns (114) (see also CARBON AND ARTIFICIAL GRAPHITE).

Table 6. Charcoal Production for Selected Years ^a

Year	Production, 10 ³ t
1940	227
1950	227
1960	290
1970 ^b	454
1980 ^b	726

^a Ref. 113.

^b Charcoal briquet production from all sources in-cluding wood, bark, lignite, coal, and agricultural residue.

Economic Aspects

Timber production in the United States is an important contributor to Gross Domestic Product (GDP). In 1991, timber-related activities in the United States generated ~\$59,498 million (Table 7). It accounted for 2.2° of the goods and structures portion of GDP. Primary timber products production totaled \$19,370; secondary timber-related products added \$40,128 million of value in 1991.

Table 7. Value of Primary and Secondary Timber Products and Gross Domestic Product (GDP) in the United States, 1991

Product and GDP	Value		GDP ^a	
	Million dollars	Percent of total	Percent of goods	Percent of total
		<i>Timber</i>		
primary	19,370	32.6	0.7	0.3
secondary	40,128	67.4	1.5	0.7
<i>Total</i>	<i>59,498</i>	<i>100.0</i>	<i>2.2</i>	<i>1.0</i>
		<i>GDP</i>		
goods ^b	2,690,200	47.0		
services	3,032,700	53.0		
<i>Total</i>	<i>5,722,900</i>	<i>100.0</i>		

^a Ref. 115.

^b Includes structures.

Primary Timber Products. Primary timber products are roundwood products such as "logs, bolts, and other round timber generated from harvesting trees for industrial or consumer use" (116). In 1991, 17,889 million ft³ of roundwood timber products were harvested in the United States (Table 8) (116). Nearly one-half (48^o %) of this volume originated in one region, the South (Table 9). Just under one-fourth (23^o %) originated from each of two regions, the North and Pacific Coast regions. Overall, 63^o % of the volume harvested was softwood, 37^o % hardwood. Ninety percent or more of the timber harvested from the Rocky Mountain and Pacific Coast regions was softwood species. Nearly two-thirds (64^o %) of the timber harvested in the South was softwood, whereas in the North, less than one-fourth (22^o %) was softwood.

Table 8. Volume and Value of Roundwood Timber Products Harvested in the United States by Region, 1986 and 1991^a

Region	Volume				Value ^b			
	Total, million ft ³	Total, %	Soft-wood, %	Hard-wood, %	Total, million dollars	Total, %	Soft-wood, %	Hard-wood, %
1986								
North	4,079	23	22	78	1,627	13	23	77
South	8,079	46	66	34	5,110	40	77	23
Rocky Mountain	948	5	90	10	821	6	94	6
Pacific Coast	4,486	26	96	4	5,082	40	98	2
United States	17,593	100	64	36	12,640	100	79	21
1991								
North	4,140	23	22	78	2,746	14	20	80
South	8,613	48	64	36	7,976	41	71	29
Rocky Mountain	938	5	90	10	1,320	7	95	5
Pacific Coast	4,198	23	93	7	7,328	38	97	3
United States	17,889	100	63	37	19,370	100	76	24

^a Refs. 116–118.

^b Market values at local points of delivery.

Table 9. Regions of the United States

North	South	Rocky Mountain	Pacific Coast
Connecticut	Alabama	Arizona	Alaska
Delaware	Arkansas	Colorado	California
Illinois	Florida	Idaho	Hawaii
Indiana	Georgia	Kansas	Oregon
Iowa	Kentucky	Nebraska	Washington
Maine	Louisiana	Nevada	
Maryland	Mississippi	New Mexico	
Massachusetts	North Carolina	North Dakota	
Michigan	Oklahoma	Montana	
Minnesota	South Carolina	South Dakota	
Missouri	Tennessee	Utah	
New Hampshire	Texas	Wyoming	
New Jersey			
New York			
Ohio			
Pennsylvania			
Rhode Island			
Vermont			
Virginia			
West Virginia			
Wisconsin			

Total roundwood harvested increased by less than 300 million ft³ between 1986 and 1991 (Table 8). The value of all roundwood timber harvested in the United States in 1991 was estimated to be \$19,370 million (Table 8). The value of roundwood timber is defined as the market value at local points of delivery, ie, delivered to a processing facility.

Softwood roundwood tended to be relatively higher valued than hardwood in all regions except the North in 1986 and 1991. Overall, softwood roundwood accounted for 63^o % of total production in 1991 but accounted for 76^o % of total value (Table 8).

Total value of roundwood production increased by \$6,730 million between 1986 and 1991 even though production increased by less than 300 million ft³ (118) (Table 8). Much of this increase is directly attributable to rapidly rising stumpage prices.

Secondary Timber Products. Secondary timber products are products manufactured from primary timber products. Secondary products can be sold directly to the final consumer or can require additional processing before reaching the final consumer. The wide diversity of products manufactured from primary timber products makes it difficult to precisely define secondary products. Lumber, for example, is clearly a secondary product because it is manufactured from roundwood and typically requires further processing before reaching its final use. Wooden furniture is considered a final product, not a secondary product because it is made from lumber or other secondary timber products. In general, products made from secondary timber products were not included in this analysis.

Table 10 itemizes the specific four-digit Standard Industrial Classification (SIC) industries considered to be secondary timber products manufacturers. For more information on the SIC system, see Executive Office of the President, Office of Management and Budget (119).

Table 10. Four-Digit SIC Secondary Timber-Related Manufacturing Industries in the United States^a

SIC code	Description
----------	-------------

24--	Lumber and Wood Products
242-	sawmills and planing mills
2421	sawmills and planing mills, general
2426	hardwood dimension and flooring mills
2429	special product sawmills, nec
243-	millwork, plywood, and structural members
2435	hardwood veneer and plywood
2436	softwood veneer and plywood
2439	structural wood members, nec
244-	wood containers
2441	nailed wood boxes and shook
2448	wood pallets and skids
2449	wood containers, nec
249-	miscellaneous wood products
2491	wood preserving
2493	reconstituted wood products
2499	wood products, nec
26--	Paper and Allied Products
261-	pulp mills
2611	pulp mills
262-	paper mills
2621	paper mills
263-	paperboard mills
2631	paperboard mills
28--	Chemicals and Allied Products
286-	industrial organic chemicals
2861	gum and wood chemicals

^a Ref. 119.

In 1991, timber-related secondary products manufacturing industries added an estimated \$40,128 million of value to primary timber products (Table 11). Most of the timber-related value added (63%) originated in the paper and allied products industry. The lumber and wood products industry added nearly 37% of total timber-related value added. Less than 1% was from chemicals and allied products.

Table 11. Value Added by Manufacture^a for Timber-Related Two-Digit SIC Industries^b by Region, 1991^c

Region	Value, 10 \$			
	Principal Group 24 (Lumber and wood products)	Principal Group 26 (Paper and allied products)	Principal Group 28 (Chemicals and allied products)	Total
North	4,067	11,385	97	15,549
South	5,367	10,846	128	16,350
Rocky Mountain	908	169	7	1,083
Pacific Coast	4,386	2,753	7	7,145
United States	14,737	25,153	238	40,128

^a Value of industry shipments minus the cost of materials, supplies, containers, fuel, purchased electricity, and contract work.

^b Timber-related industries include the following four-digit SIC Codes: 24: 2421, 2426, 2429, 2435, 2436, 2439, 2441, 2448, 2449, 2491, 2493, 2499 26: 2611, 2621, 2631 28: 2861

^c Ref. 120.

The South and North were the two largest timber-related secondary products manufacturing regions, adding \$16,350 million (41%) and \$15,549 million (30%) of value, respectively, in 1991 (Table 10). The Rocky Mountain and Pacific Coast regions combined added the remaining 20% .

BIBLIOGRAPHY

"Wood" in *ECT* 1st ed., Vol. 15, pp. 72–102, by E. G. Locke, R. H. Baechler, E. Beglinger, H. D. Bruce, J. T. Drow, K. G. Johnson, D. G. Laughnan, B. H. Paul, R. C. Rietz, J. F. Saeman, and H. Tarkow, U.S. Department of Agriculture; "Wood" in *ECT* 2nd ed., Vol. 22, pp. 358–387, by H. Tarkow, A. J. Baker, H. W. Eickner, W. E. Eslyn, G. J. Hajny, R. A. Hahn, R. C. Koeppen, M. A. Millet, and W. E. Moore, U.S. Department of Agriculture; "Wood" in *ECT* 3rd ed., Vol. 24, pp. 579–611 by Theodore H. Wegner, A. J. Baker, B. A. Bendtsen, J. J. Brenden, W. E. Eslyn, J. F. Harris, J. L. Howard, R. B. Miller, R. C. Pettersen, J. W. Rowe, R. M. Rowell, W. T. Simpson, D. F. Zinkel, U.S. Department of Agriculture.

1. *1996 Statistics, Data through 1995, Paper, Paperboard & Wood Pulp*, American Forest & Paper Association, Washington, D.C., 1996.
2. *An Analysis of the Timber Situation in the United States: 1989–2040*, General Technical Report RM-199, U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experimental Station, Fort Collins, Colo., 1990.
3. *Tree Biomass—A State-of-the-Art Compilation*, General Technical Report WO-33, U.S. Department of Agriculture, Forest Service, Washington, D.C., 1981.
4. J. L. Howard, *U.S. Production, Trade, Consumption, and Price Statistics 1965–1994*, General Technical Report FPL-GTR-98, U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, Wisc., 1997.
5. A. J. Panshin and C. deZeeuw, *Textbook of Wood Technology: Structure, Identification, Uses, and Properties of the Commercial Woods of the United States*, 4th ed., McGraw-Hill Book Co., Inc., New York, 1980.
6. *Wood Handbook: Wood as an Engineering Material*, Agriculture Handbook 72, U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, Wisc., revised, 1987.
7. R. L. Gray and R. A. Parham, *Chemtech* **12**(4), 232 (1982).
8. E. Sj nstr m, *Wood Chemistry: Fundamentals and Applications*, 2nd ed., Academic Press, San Diego, Calif., 1992.
9. D. Fengel and G. Wegener, *Wood: Chemistry, Ultrastructure, Reactions*, W. de Gruyter, Berlin and New York, 1984.
10. R. Pettersen, "The Chemical Composition of Wood," in R. Rowell, ed., *The Chemistry of Solid Wood*, American Chemical Society, Washington, D.C., 1984, Chapt. 2.
11. J. W. Rowe, ed., *Natural Products of Woody Plants. Chemicals Extraneous to the Lignocellulosic Cell Wall*, Vol. I and II, Springer-Verlag, Berlin, 1989.

12. C. Skarr and W. T. Simpson, *For. Prod. J.* **18**, 49 (1968).
13. A. J. Stamm, *Wood and Cellulose Science*, Ronald Press Company, New York, 1964.
14. F. F. P. Kollmann and W. A. Cödtü, Jr., *Principles of Wood Science and Technology*, Vol. 1, Springer-Verlag New York, Inc., New York, 1968.
15. R. C. Rietz and R. H. Page, *Air Drying of Lumber: A Guide to Industry Practices*, Agriculture Handbook 402, U.S. Department of Agriculture, U.S. Government Printing Office, Washington, D.C., 1971.
16. E. F. Rasmussen, *Dry Kiln Operators Manual*, Agriculture Handbook 188, U.S. Department of Agriculture, U.S. Government Printing Office, Washington, D.C., 1961.
17. *National Design Specifications for Wood Construction*, National Forest Products Association, Washington, D.C., 1977.
18. C. C. Gerhards, *Wood Fiber* **14**(1), 4 (1982).
19. D. E. Kretchmann and D. W. Green, *Wood and Fiber Science* **28**(3), 320–337 (1996).
20. D. W. Green and J. W. Evans, "Moisture Content and the Mechanical Properties of Dimension Lumber: Decisions for the Future" in D. Green and co-eds., *Proceedings no. 47363, In-Grade Testing of Structural Lumber*, Madison, Wisc., April 25–26, 1988, Forest Products Society, Madison, Wisc., 1989.
21. C. C. Gerhards and J. M. McMillen, eds., *Proceedings of the Research Conference on High-Temperature Drying Effects on Mechanical Properties of Softwood Lumber*, U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, Wisc., 1976.
22. H. D. Erickson and L. W. Rees, *J. Agric. Res. (Washington, D.C.)* **60**, 593 (1940).
23. W. S. Thompson, *Effect of Chemicals, Chemical Atmospheres, and Contact with Metals on Southern Pine: A Review*, Research Report 6, Forest Products Utilization Laboratories, Mississippi State University, State College, Miss., 1969.
24. *AWPA Book of Standards*, American Wood-Preservers' Association, Woodstock, Md., 1997.
25. J. E. Winandy, "Effects of Waterborne Preservative Treatment on Mechanical Properties: A Review," in *Proceedings of the 91st Annual Meeting of the American Wood-Preservers' Association*, New York, May 21–24, 1995, Vol. 91, AWWA, Woodstock, Md., 1995, pp. 17–33.
26. S. L. LeVan and J. E. Winandy, *Wood and Fiber Science* **22**(1), 113–131 (1990).
27. J. E. Winandy and P. K. Lebow, *Wood and Fiber Science* **28**(1), 39–52 (1996).
28. F. Shafizadeh and A. G. W. Bradbury, *J. Thermal Insulation* **2**, 141 (Jan. 1979).
29. *Ignition and Charring Temperature of Wood*, Report 1464, U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, Wisc., 1958.
30. A. F. Matson, R. E. Dufour, and J. F. Breen, *Survey of Available Information on Ignition of Wood Exposed to Moderately Elevated Temperatures*, Bulletin of Research 51, Part II, Underwriters Laboratories, Inc., 1959.
31. E. L. Schaffer, *Fire Technol.* **16**(1), 22 (Feb. 1980).
32. V. W. Jahnsen, Ph.D. thesis, Purdue University, West Lafayette, Ind., 1961.
33. R. W. Porter, Ph.D. thesis, Michigan State University, East Lansing, Mich., 1963.
34. J. J. Brenden, *Effect of Fire-Retardant and Other Inorganic Salts on Pyrolysis Products of Ponderosa Pine at 250° and 350°C*, Research Paper FPL 80, U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, Wisc., 1967.
35. A. J. Stamm, *Ind. Eng. Chem.* **48**, 413 (1956).
36. H. W. Eickner and W. K. Tang, *Proceedings of International Union of Forest Research Organizations*, Section 41, U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, Wisc., 1963.
37. F. L. Browne, *Theories of the Combustion of Wood and its Control*, Report 2136, U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, Wisc., 1958; reviewed and reaffirmed 1963.
38. F. C. Beall and H. W. Eickner, *Thermal Degradation of Wood Components: A Review of the Literature*, Research Paper FPL 130, U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, Wisc., 1970.
39. I. S. Goldstein in D. D. Nicholas, ed., *Degradation and Protection of Wood*, Vol. 1, Syracuse University Press, Syracuse, N.Y., 1973.
40. F. Shafizadeh, *Adv. Carbohydr. Chem.* **23**, 419 (1968).
41. *Standard Test Method for Behavior of Materials in a Vertical Tube Furnace at 750°C*, Standard E 136, American Society for Testing and Materials, West Conshohocken, Pa., 1994.
42. *Standard Methods of Fire Tests of Building Construction and Materials*, Standard E 119, American Society for Testing and Materials, West Conshohocken, Pa., 1995.
43. *Standard Methods of Fire Tests of Door Assemblies*, Standard E 252, National Fire Protection Association, Quincy, Mass., 1995.
44. *Standard Test Method for Surface Burning Characteristics of Building Materials*, Standard E 84, American Society for Testing and Materials, West Conshohocken, Pa., 1995.
45. E. L. Schaffer, *Review of Information Relating to the Charring Rate of Wood*, Research Note FPL-0145, U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, Wisc., 1966.
46. R. H. White and E. V. Nordheim, *Fire Technol.* **28**, 5 (1992).
47. R. H. White, in P. J. DiNunno and co-eds., *The SFPE Handbook of Fire Protection Engineering*, 2nd ed., National Fire Protection Association, Quincy, Mass., 1995, Chapt. 4, p. 217–229.
48. H. W. Eickner, *ASTM Journal of Materials* **1**, 625 (1966).
49. J. J. Brenden, *Rate of Heat Release from Wood-base Building Materials Exposed to Fire*, Research Paper FPL230, U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, Wisc., 1974.
50. H. C. Tran, in V. Babrauskas and S. J. Grayson, eds., *Heat Release in Fires*, Elsevier Applied Science, New York, 1993, Chapt. 11, Part b, pp. 357–372.
51. *Standard Test Method for Heat and Visible Smoke Release Rates for Materials and Products Using an Oxygen Consumption Calorimeter* Standard E 1354, American Society for Testing and Materials, West Conshohocken, Pa., 1994.
52. J. E. Winandy, R. J. Ross, and S. L. LeVan, in *Proceedings of the 1991 International Timber Engineering Conference*, 1991 Sept. 2–5, Vol. 4, London, TRADA, London, 1991, pp. 4.69–4.74.
53. S. C. Juneja and L. Calve, *J. Fire Retardant Chem.* **4**, 235–241 (1977).
54. S. L. LeVan, *Effectiveness of Fire-Retardant Treatments for Shingles after 10 years of Outdoor Weathering*, Research Paper FPL 474, U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, Wisc., 1986.
55. R. S. Roberts, *Biotechnol. Bioeng. Symp.* **10**, 125 (1980).
56. R. H. Baechler, *J. For. Prod. Res. Soc.* **4**, 332 (1954).
57. R. H. Baechler and C. A. Richards, *Trans. Amer. Soc. Mech. Eng.* **73**, 1055 (1951).
58. W. E. Eslyn and J. W. Clark, *Mater. Org. Suppl.* **3**, 43 (1976).
59. G. R. Esenther and R. H. Beal, *Sociobiology* **4**, 215 (1979).
60. *Comparative Decay Resistance of Heartwood of Native Species*, Research Note FPL 0153, U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, Wisc., 1967.
61. *Wood Preservation: Treating Practices*, Federal Specification TT-W-5711, U.S. General Services Administration, Washington, D.C.
62. T. C. Scheffer and W. E. Eslyn, *For. Prod. J.* **28**, 25 (1978).
63. T. L. Highley, *For. Prod. J.* **30**, 49 (1980).
64. E. Schmidt, *First International Conference on Wood Protection with Diffusible Preservatives*, 1990, pp. 91–92.

65. J. A. Meyer, *Wood Sci.* **14**(2), 49 (1981).
66. R. M. Rowell and W. B. Banks, *Water Repellency and Dimensional Stability of Wood*, General Technical Report FPL 50, U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, Wisc., 1985.
67. R. M. Rowell and R. L. Youngs, *Dimensional Stability of Wood in Use*, Research Note FPL 0243, U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, Wisc., 1981.
68. R. M. Rowell, *Proc. Am. Wood-Preserv. Assoc.* **71**, 41 (1975).
69. E. C. O. Erikson, *Mechanical Properties of Laminated Modified Wood*, Report 1639, U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, Wisc., 1965.
70. L. F. Gjovik and H. L. Davidson, *Comparison of Wood Preservatives in Stake Tests*, Research Note FPL 02, U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, Wisc., 1979.
71. R. M. Seborg and A. E. Vallier, *J. For. Prod. Res. Soc.* **4**, 305 (1954).
72. C. Schuerch, *For. Prod. J.* **14**, 377 (1964).
73. P. J. Ince, *Recycling of Wood and Paper Products in the United States*, General Technical Report FPL-GTR-89, U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, Wisc., 1996.
74. Ref. 9, p. 338.
75. F. W. Herrick, and H. L. Hergert, "Utilization of Chemicals from Wood: Retrospect and Prospect," in F. A. Loewus and V. C. Runeckles, eds., *The Structure, Biosyntheses, and Degradation of Wood, Recent Advances in Phytochemistry*, Vol. 11, Plenum Press, New York, 1977.
76. *Annual Pulpwood Statistics Summary Report*, American Pulpwood Association Inc., Rockville, Md., 1995.
77. D. Zinkel and J. Russell, eds., *Naval Stores: Production, Chemistry, Utilization*, Pulp Chemicals Association, New York, 1989.
78. *Naval Stores Review*, New Orleans, La., 1995.
79. J. F. Harris, J. F. Saeman, and E. G. Locke inn B. L. Browning, ed., *Wood as a Chemical Raw Material, The Chemistry of Wood*, Interscience Publishers, a division of John Wiley & Sons, Inc., New York, 1973, Chapt. 11, pp. 535–585.
80. J. A. Hall, J. F. Saeman, and J. F. Harris, *Unasyhva* **10**(1), 7 (1956).
81. J. F. Harris, J. F. Saeman, and E. G. Locke, *For. Prod. J.* **8**, 248 (1958).
82. G. J. Hajny, *Biological Utilization of Wood for Production of Chemicals and Foodstuffs*, Research Paper FPL 385, U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, Wisc., 1981.
83. D. F. Root and co-workers, *For. Prod. J.* **9**, 158 (1959).
84. S. W. McKibbins and co-workers, *For. Prod. J.* **12**, 17 (1962).
85. J. F. Harris and J. M. Smuk, *For. Prod. J.* **11**, 303 (1961).
86. M. S. Feather and J. F. Harris, *Adv. Carbohydr. Chem. Biochem.* **28**, 161 (1973).
87. *Chem. Eng. News*, **35** (Nov. 3, 1980).
88. J. F. Saeman, E. G. Locke, and G. K. Dickerman, *Production of Wood Sugar in Germany and its Conversion to Yeast and Alcohol*, U.S. Department of Commerce Office Technical Service PB Report 7736, U.S. Department of Commerce, Washington, D.C., 1945.
89. E. E. Harris, *Adv. Carbohydr. Chem.* **4**, 154 (1949).
90. R. A. Lloyd and J. F. Harris, *Wood Hydrolysis for Sugar Production*, Report 2029, U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, Wisc., 1955.
91. A. J. Panshin, *Wood Saccharification in Forest Products*, McGraw-Hill Book Co., Inc., New York, 1950, pp. 403–412.
92. K. N. Cederquist, *Some Remarks on Wood Hydrolyzation, Report of a Seminar Held at Lucknow, India, 1952, on Production and Use of Power Alcohol in Asia and the Far East*, Technical Assistance Administration, United Nations, New York, 1954, pp. 193–198.
93. J. D. Broder, J. W. Barrier, K. P. Lee, and M. M. Bulls, *World Resources Review* **7**(4), (1995).
94. A. J. Baker and T. W. Jeffries, *Status of Wood Hydrolysis for Ethanol Production*, Report of U.S. Department of Agriculture, Forest Service, for U.S. Agency for International Development Support Bureau, Office of Energy (TMR Authorization No. 81–89), Washington, D.C., June 1981.
95. **U.S. Pat. 4,201,596** (May 6, 1980), J. A. Church and co-workers (to American Can Co.).
96. R. D. Ziminski, *paper presented at American Institute of Chemical Engineers Meeting, New Orleans, LA*, Nov. 1981.
97. D. R. Thompson and H. E. Grethlein, *Ind. Eng. Chem. Prod. Res. Dev.* **18**(3), 166 (1979).
98. G. T. Tsao and co-workers, *Lorrie Biomass Conversion Conference*, Purdue University, West Lafayette, Ind., 1981.
99. R. S. Roberts, *Biotechnol. Bioeng. Symp.* **10**, 125 (1980).
100. *Enzymatic Hydrolysis of Cellulose to Glucose—A Report on the Natick Program*, U.S. Army, Natick Research and Development Command, Natick, Mass., Sept. 1981.
101. L. Spano, *J. Coatings Technol.* **50**, 71 (1978).
102. G. H. Emert, R. Katzen, R. E. Fredrickson, and K. F. Kaupisch, *Chem. Eng. Prog.* **76**(9), 47 (1980).
103. D. M. Jenkins and T. S. Reddy, *Economic Evaluation of the MIT Process for Manufacture of Ethanol*, DSE-3992-T1, National Technical Information Service, Washington, D.C., 1979.
104. R. Brooks, T.-M. Su, M. Brennan, and J. Frick, *Proceedings 3rd Annual Biomass Energy Systems Conference, Golden, CO*, SERI TP-33-285, National Technical Information Service, Washington, D.C., June 1979.
105. R. H. Marchessault, S. Coulombe, T. Hanai, and H. Morikawa, *Monomers and Oligomers from Wood*, *Transactions of Technical Section of the TR52*, 1980; published in *Pulp and Paper Canada* **81**(6), 1980.
106. L. Jurasek, *Devel. Ind. Microbiol.* **1978**, **177** (1979).
107. L. A. Underkofler and R. J. Hickey, *Industrial Fermentation*, Chemical Publishing Company, Inc., New York, 1954.
108. H. D. Turner, *For. Prod. J.* **14**, 282 (1964).
109. J. F. Harris, *Tappi* **61**, 41 (1978).
110. **U.S. Pat. 4,280,878** (Oct. 30, 1979), G. E. Sprenger.
111. J. Rienks, *Forest Products Research Society Proceedings No. P-75-13*, Forest Products Research Society, Madison, Wisc., pp. 104–106, 1975.
112. J. Hartwig, *For. Prod. J.* **21** (1971).
113. A. J. Baker, "Charcoal," in R. C. Davis, ed., *Encyclopedia of American Forest and Conservation History*, Vol. 1, Macmillan Publishing Company, New York, 1983, pp. 73–77.
114. D. E. Earl, *A Report on Charcoal*, Food and Agriculture Organization of the United Nations, Rome, Italy, 1974.
115. *Federal Reserve Bulletin* **79**(12):A51, Board of Governors of the Federal Reserve System, Washington, D.C., 1993.
116. D. S. Powell, J. L. Faulkner, D. R. Darr, Z. Zhu, and D. W. MacCleery, *Forest Resources of the United States, 1992*, General Technical Report RM-234, U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station, Ft. Collins, Colo., 1993, 132 pp.
117. K. L. Waddell, D. D. Oswald, and Douglas S. Powell, *Forest Statistics of the United States, 1987*, Resource Bulletin PNW-RB-168, U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, Oreg., 1989, 106 pp.
118. D. B. McKeever and K. C. Jackson, "Economic Importance of the Timber Processing Industries," in R. W. Haynes, coordinator, *An. Analysis of the Timber Situation in the United States, 1989–2040*, General Technical Report RM-199, U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colo., 1990, Chapt. 4, pp. 59–78.
119. Executive Office of the President, Office of Management and Budget, *Standard Industrial Classification Manual, 1987*, National Technical Information Service, Springfield, Va., 1987, 703 pp.

120. *1991 Annual Survey of Manufactures*, Geographic Area Series, 1987 Economic Censuses compact disk, Vol. 1, Release 1E, Data User Services Division, U.S. Department of Commerce, Washington, D.C., 1993.

Irene Durbak
David W. Green
Terry L. Highley
James L. Howard
David B. McKeever
Regis B. Miller
Roger C. Pettersen
Roger M. Rowell
William T. Simpson
Kenneth E. Skog
Robert H. White
Jerrold E. Winandy
John I. Zerbe
USDA Forest Service



WOOL

Wool is unique among clothing fibers: it is not only the oldest, it is also the only fiber to have been used continuously and universally. Wool is the fibrous covering from sheep (1) and is by far the most important animal fiber used in textiles. World greasy wool production was 2,688,000 tons in 1994–1995, equivalent to 1,557,000 t clean (2) (Table 1). In 1994–1995, 1000.1×10^6 sheep produced $2,688 \times 10^6$ t of greasy wool. The average clip of 2.69 kg sheep produces 1.56 kg sheep of clean wool (Tables 1 and 2).

Table 1. World Production of Wool 1994–1995^a, t

Source	Greasy	Clean
Merino	1,116,000	666,000
Crossbred	682,000	395,000
Other (carpet)	890,000	496,000
<i>Total</i>	<i>2,688,000</i>	<i>1,557,000</i>

^a Ref. 2.

Table 2. Wool Production and Numbers of Sheep in Principal Wool-Producing Countries, 1994–1995^a

Country region	Wool production (greasy)		Sheep, $\times 10^6$
	$\times 10^3$ t	$^{\circ}_o$	
Australia	724	26.9	120.4
New Zealand	289	10.8	48.6
South Africa	68	2.5	18.5 (1993 94)
Argentina	90	3.3	21.2
Uruguay	88	3.3	21.8
United States	32	1.2	8.9
United Kingdom	66	2.5	29.5
Turkey	74	2.8	39.5
Former Soviet Union, China, Eastern Europe	669	24.9	236 ^b
other	588	21.9	455.7
<i>Total</i>	<i>2688^c</i>	<i>100</i>	<i>1000.1^b</i>

^a Ref. 2.

^b Value is estimated.

^c Clean equivalent is 1539×10^3 t

Wool belongs to a family of proteins, the keratins, that also includes hair and other types of animal protective tissues such as horn, nails, feathers, and the outer skin layers. The relative importance of wool as a textile fiber has declined over the decades as synthetic fibers have increasingly been used in textile consumption. Wool is still an important fiber in the middle and upper price ranges of the textile market. It is also an extremely important export for several nations, notably Australia, New Zealand, South Africa, and Argentina and commands a price premium over most other fibers because of its outstanding natural properties of soft handle (the feel of the fabric), moisture absorption abilities (and hence comfort), and superior drape (the way the fabric hangs) (see FIBERS; TEXTILES). Table 2 shows wool production and sheep numbers in the world's principal wool-producing countries.

The principal characteristics of clean wool types are average diameter, measured in micrometers, and average length, measured in millimeters (Table 3).

Table 3. Main Types of Wool

Wool type	Breed examples	Fiber diameter, μm	Fiber length, mm
fine	merino	10–30	35–100
	cheviot	20–40	50–100
medium	dorset		
	hampshire		
	southdown		
	columbia	20–30	75–150
	corriedale		
crossbred	polwarth		
	targhee		
	cotswold	25–50	125–350
	leicester		
	lincoln		

The finest diameter wool is produced by merino sheep, an animal first bred in Spain but now prevalent throughout the world. Over 75% of the sheep in Australia, the world's largest wool producer, are merino sheep, which are also bred in large numbers in South Africa, Argentina, and the former USSR.

Medium diameter wool includes sheep breeds of English origin, eg, southdown, hampshire, dorset, and cheviot, as well as crossbreds, eg, columbia, targhee, corriedale, and polwarth, from interbreeding with merinos. Coarse diameter wool comes from sheep chiefly bred for mutton, eg, lincoln, cotswold, and leicester.

Raw wool from sheep contains other constituents considered contaminants by wool processors. These can vary in content according to breed, nutrition, environment, and position of the wool on the sheep. The main contaminants are a solvent-soluble fraction called wool grease; protein material; a water-soluble fraction (largely perspiration salts collectively termed suint); dirt; and vegetable matter, eg, burrs and seeds from pastures. Table 4 gives some figures for percentages of fleece constituents in Australian wool, excluding the protein contaminant (3,4). The wools in Table 4 contain little or no vegetable matter, which in certain areas can be as high as 20% of the raw-wool weight, but is usually <2%.

Table 4. Fleece Composition of Australian Wool, %

Component	Sheep			Lambs, Merino, and Crossbred
	Merino	Crossbred	Skin ^a	
oven-dry wool				
max	66.8	72.2	70.0	68.0
min	29.4	49.3	50.6	54.0
av	48.9	61.0	63.0	60.5
wax				
max	25.4	19.3	20.0	23.5
min	10.0	5.3	9.9	6.4
av	16.1	10.6	15.8	16.0
suint				
max	13.0	13.6	1.4	7.2
min	2.0	4.4	0.1	3.4
av	6.1	8.2	0.6	5.4
dirt				
max	43.8	23.7	21.5	10.0
min	6.3	4.3	6.4	3.8
av	19.6	8.4	11.2	6.4
moisture				
max	12.6	14.2	9.6	14.2
min	8.1	9.5	7.2	9.0
av	9.6	12.0	8.0	11.2

^a Skin wool is the term applied to fellmongered wool, usually removed from skins by a sweating process involving bacterial action or by action of lime-sulfide depilatory.

Raw-Wool Specification

In buying raw wool, wool processors are concerned about its quality and the quantity of pure fiber present. The particular properties of concern differ depending on the processing system used and the product. For example, for the main wool market, ie, fine and medium wools for apparel, the single-fiber and staple characteristics that are significant to processing are given in Table 5 (5). The most important are the yield, ie, the percentage content of pure fiber at a standard (usually 16% moisture content; the vegetable-matter content and type; the average fiber diameter; the average length of the fibers; the strength of the staples of fibers and the position of any weak spot along the fibers; the color of the clean wool; and the number (if any) of naturally colored fibers present. For carpet-type wools (long wools) the important properties (6) are yield; fiber diameter; fiber length; color; bulk (the volume occupied by the fibers in a yarn); medullation (see FIBERS); and vegetable-matter content.

Table 5. Significance of Single-Fiber and Staple Characteristics in Wool Textile Processing^a

Characteristic	Importance
yield	++++
fiber diameter	++++
vegetable matter	+++
staple length	+++
staple strength	+++
color	++
dark fibers	++
resistance to compression	+
style (eg, tip, crimp, dust)	+
handle	+

^a Ref. 5.

Until the late 1960s and early 1970s, the characteristics of different wools (see Table 5) were largely evaluated visually by wool classers and valuers, with all the inherent faults implied by a subjective measuring system. With the development of sampling techniques and equipment capable of rapid and economical measurement of yield, diameter, and vegetable matter (7,8), the sale of wool by sample and objective measurement has become dominant in major wool exporting countries. In sale by sample, cores are drawn from each lot and tested for yield, diameter, and vegetable-matter content in accordance with international standards (9). In addition, a full-length display sample, representative of each lot and obtained by standard procedure (10), is available for buyers to appraise the other characteristics. Sale by sample has reduced costs by reducing the handling of bulk wool in wool-brokers' stores and selling operations, and has also led to more efficient utilization of wool in processing and more accurate alerting of woolgrowers to wool processors' requirements.

Techniques for efficiently and economically measuring the other important characteristics, ie, length, strength, and position of weakness in the fiber, are now in commercial use. Existing color-measuring equipment can be used to measure the color (whiteness or yellowness) of washed wool, but accurate measuring of colored-fiber content remains a problem (5,8).

The overall objective of research under way as of ca 1997 is to develop a system of sale by description for fine and medium wools whereby the buyer is presented only with measured data on the principal characteristics of the raw wool, as well as an assessment of the less important characteristics by an independent skilled appraiser (8). A scheme for assessing the risk of the presence of colored fiber content in greasy wool has been proposed which depends on production parameters and on the age and sex of the sheep (5). Instrumentation and computer algorithms for the measurement of style and handle

parameters related to crimp frequency and definition, staple shape, dust penetration, and raw wool color have also been developed (5). The overall objective of these developments and of continuing research is to facilitate the introduction of a system of sale of raw wool by description in which a sale sample will not be required for inspection.

Fiber Growth

Wool follicles are produced from multicellular tube-like structures known as follicles. These follicles are located in the skin layers (dermis and epidermis) of sheep, and two types of follicles, primary and secondary, are usually identified. Primary and secondary follicles are described from the order of their initiation and development in fetal skin. The primary follicles develop first, in the unborn lamb. Secondary follicles develop later, and in finer woolled sheep derived secondary follicles subsequently form by branching from the original secondaries with which they share a common orifice (11). Each primary follicle has a sebaceous gland and a sweat gland together with an arrector muscle, whereas secondary follicles usually have only an associated sebaceous gland (12).

Fiber Morphology

Wool fibers consist of cells, where flattened overlapping cuticle cells form a protective sheath around cortical cells. In some coarser fibers, a central vacuolated medullary cell type may be present.

In fine wool such as that obtained from merino sheep, the cuticle is normally one cell thick ($20 \times 30 \times 0.5$ mm, approximate dimensions) and usually constitutes about 10% by weight of the total fiber. Sections of cuticle cells show an internal series of laminations (Figs. 1 and 2) comprising outer sulfur-rich bands known as the exocuticle and inner regions of lower sulfur content called the endocuticle (13). On the exposed surface of cuticle cells, a membrane-like proteinaceous band (epicuticle) and a unique lipid component form a hydrophobic resistant barrier (14). These lipid and protein components are the functional moieties of the fiber surface and are important in fiber protection and textile processing (15).

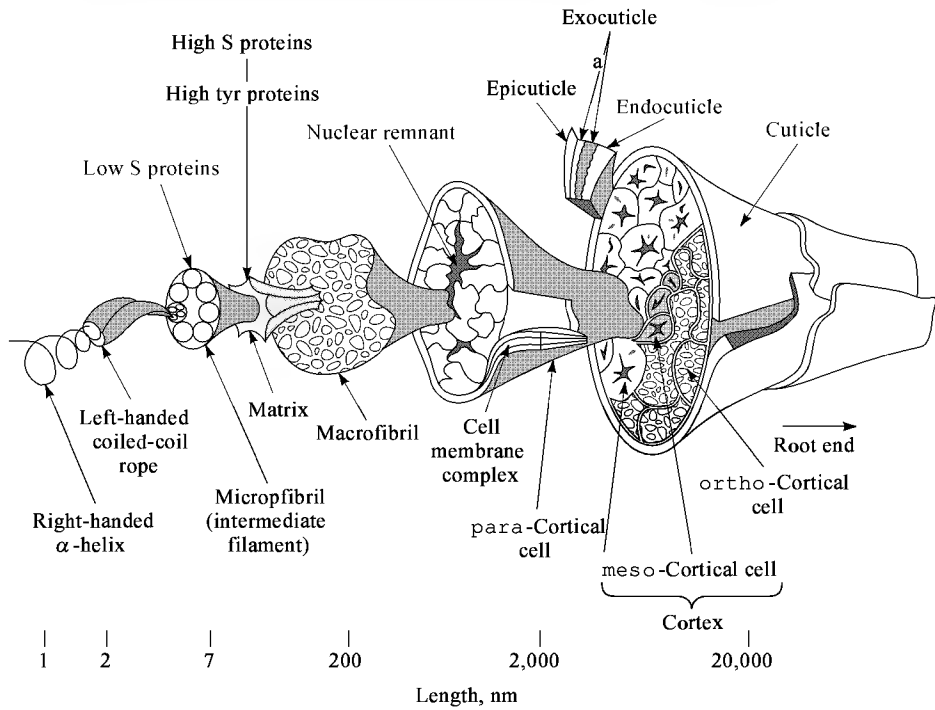


Fig. 1. Schematic of the structure of a fine Merino wool fiber.

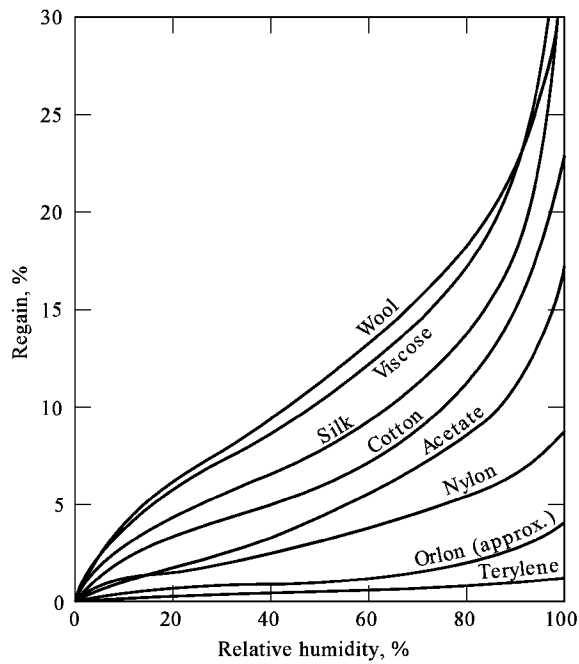


Fig. 2. The moisture sorption isotherm of various fibers.

The cortex comprises the main bulk and determines many mechanical properties of wool fibers (see Fig. 1). Cortical cells are long, polyhedral, and

spindle-shaped (approximately 100 nm long) and consist of a class of biological filaments known as intermediate filaments (16) embedded in a sulfur-rich matrix. The intermediate filaments (originally termed microfibrils) together with the matrix are organized into large macrofibrillar units and these are often observed in sections of cortical cells. In fine wool fiber sections obtained from merino wool, the main cortices known as ortho- and para- are arranged bilaterally. Orthocortical cells show different intermediate filament matrix packing arrangements from those of paracortical cells (17).

The arrangement of ortho- and paracortical cells differs among wool types. For example, in lincoln wool an annular (orthocortical core surrounded by paracortex) cellular arrangement is present. Merino fibers possess a characteristic crimp and in these fibers the orthocortex is located on the outer side of the crimp curvature.

Between cuticle cells and cortical cells a continuous intercellular material is present which, despite being a relatively minor fraction of the total fiber weight, is a material of increasing interest owing to its presumed role in water and reagent-penetration of fibers. This region of the fiber is known as the cell membrane complex (18) and comprises the intercellular material together with the apposing cellular membranes of cuticle and cortical cells, which are approximately 25 nm thick.

Physical Properties

Moisture Sorption (Regain). One of wool's major attributes as a textile fiber is its ability to absorb and desorb large amounts of moisture as the relative humidity surrounding the fiber changes. The amount of moisture absorbed by wool is expressed as a percentage of the dry weight of wool. A pronounced hysteresis is observed in the sorption isotherm of wool, with a moisture level 2% higher on desorption than absorption at most relative humidities (19). The saturated moisture level of wool is about 33%; this is higher than that of most other fibers (see Fig. 2). The absorption of water by wool results in improved comfort during wear, with the heat of absorption buffering the wearer against sudden environmental changes; during active sports, perspiration transport away from the skin is enhanced. As can be seen from Table 6, most physical properties of the fiber are affected by the moisture level. Wool is sold by weight; thus, allowance for uptake of water must be made. A premium is paid for fine wools and since the diameter changes with moisture levels (wet to dry), diameter measurements must be made under standard conditions of temperature and humidity.

Table 6. The Effect of Moisture Content on the Physical Properties of Wool Fibers at 25°C^a

Property	Moisture content, %							
	0	5	10	15	20	25	30	33
relative humidity, % (absorption)	0	15	42	68	85	94	99	100
relative humidity, % (desorption)	0	8	32	58	79	92	98	100
specific gravity, kg m ⁻³	1.304	1.314	1.315	1.313	1.304	1.292	1.277	1.268
volume swelling, %	0	4.24	9.07	14.25	20.0	26.2	32.8	36.8
length swelling, %	0	0.55	0.93	1.08	1.15	1.17	1.18	1.19
radial swelling, %	0	1.82	4.00	6.32	8.88	11.69	14.57	16.26
heat of wetting ^b , kJ kg wool	101	64.4	38.1	20.5	10.0	4.2	1.13	0
Young's modulus (relative)	1.00	0.96	0.87	0.76	0.66	0.56	0.44	0.38
torsional rigidity modulus, GPa ^c	1.76	1.60	1.26	0.90	0.50	0.28	0.16	0.11
electrical resistivity, MΩ·m		3 × 10 ⁴	400	8	0.40	0.06		
dielectric constant at 10 ⁴ Hz	4.6	5.1	6.2	8.3	12.8			

^a Ref. 33.
^b Heat evolves when wool, dry mass of 1 kg, at a particular regain is immersed in water.
^c To convert GPa to psi, multiply by 145,000.

Diameter Distribution. Wool fibers exhibit a range of diameters from coarse fibers (25–70 μm), used in carpets, to fine merino fibers (10–25 μm), used in apparel because of their soft handle. Fibers from an individual sheep also exhibit a range of diameters. The mean diameter is generally reported in sale catalog, but the distribution of diameters is also important. In particular, when wool is worn next to the skin, the number of coarse fibers affects comfort as these fibers, rather than buckling, indent the skin and activate nerve receptors. This gives rise to a sensation of prickle and itch that has been incorrectly assumed, by some consumer, to be an allergic reaction. True allergies are rare, if they exist at all, and the irritation is mechanical and not immunological. This discomfort also applies to synthetic fibers as well as to wool and is easily prevented by selecting wools without coarse fibers. When wool was compared to a synthetic with a similar fiber diameter distribution, no difference in discomfort could be detected (20).

Felting. A unique morphological feature of the wool fiber surface is that it consists of overlapping cuticle cells. The protruding scale edges result in differential friction between the with-scale and against-scale direction. As a consequence of this, under the appropriate conditions, wool can be readily felted to produce a dense matting of fibers. These felts are then used for products as diverse as hats, polishing pads, table covers, and piano hammers and dampers. Felting shrinkage in garments is obviously undesirable but is easily prevented by reducing the differential friction. One approach involves applying a thin polymer coating to the fibers to partially mask the scale edges.

Thermal Properties. The regular packing of α-helical polypeptide chains within the intermediate filaments forms a crystalline phase that occupies about 70% of the dry volume of the fiber (21). This phase melts irreversibly at a temperature that is both time- and regain-dependent (22). During dyeing and finishing operations of wool, no melting of the fiber occurs. However, care should be taken when processing blends of wool and synthetic fibers that require higher temperatures.

The amorphous matrix phase contains a high concentration of the amino acid cystine and is therefore highly cross-linked. Similarly to other amorphous materials, a glass transition, *T_g*, has been detected in the wool fiber (23–25). Water acts as a plasticizer, lowering the glass-transition temperature of the dry fiber from 165°C to below -10 °C when wet (Fig. 3) (24). The glass-transition temperature is an important parameter, as the properties and performance of wool are influenced by the environmental conditions (temperature and humidity) relative to the glass transition. The insertion of cohesively set creases or pleats in wool fabric (equivalent to thermal set in synthetic polymers) is achieved by subjecting wool to bending strain above *T_g* and fixed by transition to conditions below *T_g* (24). This can be achieved by heating and cooling, wetting and drying or a combination of both, eg, steam pressing. Wrinkle recovery is poorer when creases are inserted above *T_g* and recovery is below *T_g* (26). This is likely to occur under hot and humid conditions where local wetting (perspiration) and crease insertion occurs simultaneously. When the fabric moves away from the skin the crease has insufficient time for complete recovery before the fabric rapidly dries causing the *T_g* of fabric to rise to a temperature above that of the environment. The viscoelastic properties (23), physical aging (23) and the water absorption isotherm of wool (25) are also influenced by the glass transition (see Table 6).

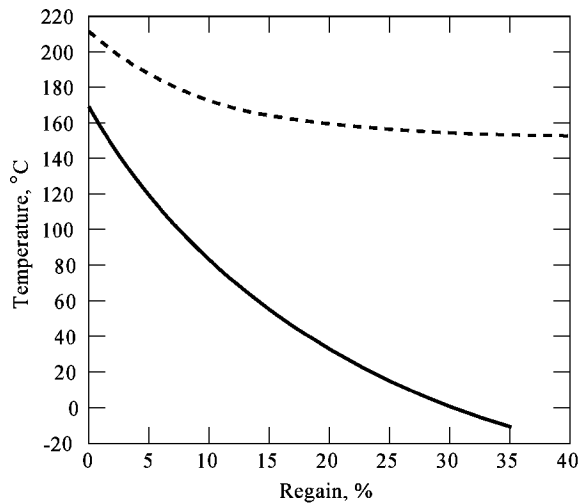


Fig. 3. The glass-transition (23) and melting temperature (22) of wool as a function of regain.

Mechanical Properties. Although wool has a complicated hierarchical structure (see Fig. 1), the mechanical properties of the fiber are largely understood in terms of a two-phase composite model (27–29). In these models, water-impenetrable crystalline regions (generally associated with the intermediate filaments) oriented parallel to the fiber axis are embedded in a water-sensitive matrix to form a semicrystalline biopolymer. The parallel arrangement of these filaments produces a fiber that is highly anisotropic. Whereas the longitudinal modulus of the fiber decreases by a factor of 3 from dry to wet, the torsional modulus, a measure of the matrix stiffness, decreases by a factor of 10 (30).

The longitudinal stress–strain curves for a wool fiber at different relative humidities are shown in Figure 4 (31). Three distinct regions can be discerned, especially for fibers at higher relative humidity. Once the fiber crimp is removed, a near-linear region up to about 2° strain is obtained (preyield region). For the wet fiber this is generally associated with stretching of the α -helices within the intermediate filaments. At lower moisture levels the matrix phase plays an increasingly dominant role. Between 2 and 25° strain (yield-region) progressive unfolding of zones of α -helices to form a β -pleat configuration occurs. Very little increase in stress is observed during this stage, whereas complete recovery is still possible provided the fiber is allowed to relax in water. Beyond 25° strain (post-yield region), the fiber stiffens and breaks. At a molecular level, the reasons for this are still a matter of debate but include resistance to the unfolding of a stabilized region of the intermediate filaments (27,29) and the rubber-like response of the matrix (28).

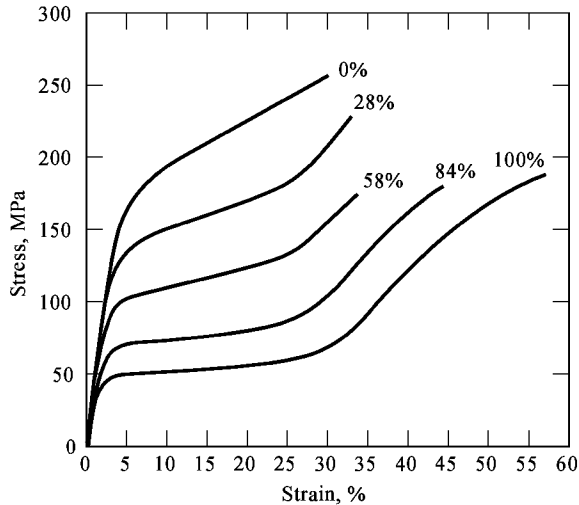


Fig. 4. The stress–strain curves of a wool fiber at different relative humidities.

For a fiber immersed in water, the ratio of the slopes of the stress–strain curve in these three regions is about 100:1:10. Whereas the apparent modulus of the fiber in the preyield region is both time- and water-dependent, the equilibrium modulus (1.4 GPa) is independent of water content and corresponds to the modulus of the crystalline phase (32). The time-, temperature-, and water-dependence can be attributed to the viscoelastic properties of the matrix phase.

Chemical Structure

Wool belongs to the family of proteins (qv) called keratins. However, morphologically the fiber is a composite and each of the components differs in chemical composition. Principally the components are proteinaceous, although wool cleaned of wax, suint, and other extraneous materials acquired during growth contains small amounts of lipids (structural and free), trace elements, and, in colored fibers, pigments called melanin.

Studies on the chemical structure of wool have been largely confined to fine merino fibers, although aspects treated herein are relevant to all wool types.

Protein Components. The simplest picture of the proteinaceous components is one of polypeptides, which are composed of α -amino acid residues. It is estimated that wool contains about 170 different types of polypeptides varying in molecular mass from below 10,000 to greater than 50,000 (34). Complete acid hydrolysis of wool yields 18 amino acids, the relative amounts of which vary considerably from one wool to another. Typical figures for two different samples of wool are given in Table 7.

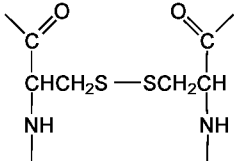
Table 7. Amino Acid Content of Merino Wool

Amino acid + H ₃ N–CH(R)–COO	Amino acid content, residues 100 residues		Nature of R group
	Bradbury and co-workers	Leeder & Marshall	
	1967	1982	
glycine	8.6	8.2	aliphatic hydrocarbon

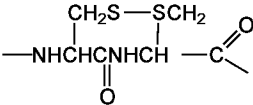
alanine	5.3	5.4	aliphatic hydrocarbon
valine	5.5	5.7	aliphatic hydrocarbon
leucine	7.7	7.7	aliphatic hydrocarbon
isoleucine	3.1	3.1	aliphatic hydrocarbon
phenylalanine	2.9	2.8	aromatic hydrocarbon
tyrosine	4.0	3.7	aromatic hydrocarbon
serine	10.3	10.5	hydroxyl
threonine	6.5	6.3	hydroxyl
aspartic acid ^a	6.4	6.6	acidic
glutamic acid ^b	11.9	11.9	acidic
hisidine	0.9	0.8	basic
arginine	6.8	6.9	basic
lysine	3.1	2.8	basic
methionine	0.5	0.4	sulfur containing
cysteine	10.5	10.0	sulfur containing
tryptophan	^c	^c	heterocyclic
proline	5.9	7.2	heterocyclic

^a Includes asparagine.
^b Includes glutamine.
^c Tryptophan is destroyed under the conditions used for these analyses; values of about 0.5 residues % have been obtained by alternative techniques (35–37).

The side groups of the amino acids vary markedly in size and chemical nature and play an important role in the chemical reactions of the fiber. For example, the basic groups (hisidine, arginine, and lysine) can attract acid (anionic) dyes, and in addition the side chains of lysine and hisidine are important sites for the attachment of reactive dyes. The sulfur-containing amino acid cysteine plays a very important role, because almost all of the cysteine residues in the fiber are linked in pairs to form cystine residues, which provide a disulfide bridge –S–S– between different polypeptide molecules or between segments of the same molecules as shown:



or



The disulfide bridges act as cross-links and prevent movement of chains and chain segments, and thus are responsible for the high form stability of wool fabrics. The disulfide cross-links are readily rearranged through a thiol–disulfide interchange reaction under the influence of heat and moisture, or more rapidly on treatment with alkaline-reducing agents. This facilitates conformational rearrangement of the wool proteins, leading to relaxation of molecular stress in the fiber. This property is employed in the permanent setting of wool fabrics.

Structure of Wool Proteins. The structure of wool proteins has been the subject of much research. Methods to solubilize, separate, and determine the amino acid sequence of these proteins have been reviewed (38–40). The proteins of wool basically belong to three groups: low sulfur proteins, rich in amino acids that contribute to α-helix formation (glutamic acid, aspartic acid, leucine, lysine, arginine); high sulfur proteins rich in cystine, proline, serine, and threonine; and high glycine–tyrosine proteins which are also rich in serine. The groups of proteins that constitute wool are not uniformly distributed throughout the fiber, but are aggregated within the various morphological regions.

The proteins derived from the rod-like microfibrils of the cortical cells contain the helical segments of low sulfur proteins set in a nonfilamentous matrix of principally high sulfur and high tyrosine proteins. It has been shown that the amino acid sequences of the helical portions are highly homologous with those of intermediate filament proteins derived from skin and other tissues, and for this reason the microfibrils of wool are now generally referred to as intermediate filaments (41). The filament proteins of the ortho- and paracortex are similar, but the proteins of highest sulfur content are concentrated in the paracortex (42). The intermediate filaments contain a rod-like central helical domain in which the sequences show a heptad repeat interrupted at three positions by short nonhelical segments. The ends of the polypeptide consist of nonhelical domains terminated in a carboxyl and N-acetyl group, respectively (38,43).

Lipids. The lipids of the fiber are believed to be constituents of the cell membranes (see FIBERS). Together with intercellular material they constitute about 1% of the fiber but play an important role in many properties, such as the intercellular diffusion of reagents. The free lipids are extractable with organic solvents and consist of fatty acids, sterols, and trace amounts of glycerides, sphinolipids, and glycolipids (44). Although phospholipid is present in the wool follicle membranes, only trace amounts of phospholipid are found in the keratinized fiber (45,46). Cholesterol and its biosynthetic precursor desmosterol are the main sterol components of the free lipids. In addition to the free lipids, wool contains some lipids that are assumed to be covalently bound to proteins. These components are not readily removed by organic solvents but can be released by mild alkaline hydrolysis. Covalently bound surface lipids account for the hydrophobicity of the fiber surface. An unusual odd-chain fatty acid, 18-methyleicosanoic acid, is the predominant fatty acid, accounting for approximately 60% of the bound fatty acid on the surface of wool.

Wool Processing

The conversion of raw wool into a textile fabric or garment involves a long series of separate processes. There are two main processing systems, ie, worsted and woollen, although an appreciable volume of wool is also processed on the short-staple (cotton) system or on the semiworsted system for carpet use. The principal stages in the woollen and worsted systems are shown in Figure 5 (47). Reference 48 provides a discussion of more recent developments in the early stages of the worsted processing sequence.

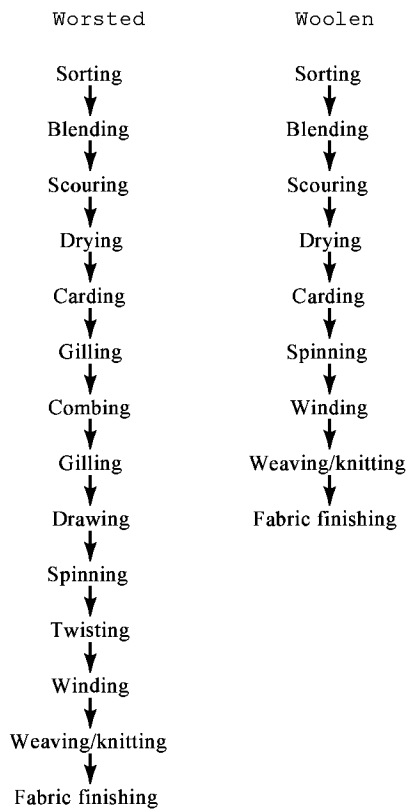


Fig. 5. The principal stages in the worsted and woollen systems.

The majority of the world's apparel wool clip is combing wool and is processed on the worsted system. This process was so named because the combed wool spinning industry in England developed in the English town of Worsted in the fourteenth century (47). Fine smooth yarns were produced by "spinsters" who placed the combed parallel array of fibers on top of a distaff (combed wool even today is known as top), drew the strand down to the required linear density, and inserted the twist using a spindle. Prior to the development of worsted spinning, and co-existing ever since with it, the traditional system produced yarns in which the fibers were much more disorientated. Although the texture was not as smooth, there was a high degree of bulk and the finished woven fabrics produced coatings, blankets, and distinctive products such as tweeds and flannels. Woollen spinning is the mechanized form of the traditional system.

Besides scoured wool, the next saleable commodity in worsted processing is top, or combed wool. Generally the combing plant combines the processes of scouring and combing. Most often, a separate commercial enterprise which buys top from the topmaker carries out spinning. Although there are woollen sales to yarn spinners, particularly in the knitwear area, the production of woven woollen products in the one enterprise from raw wool is more common. The woollen process is characterized by a much higher degree of vertical integration than the worsted system.

Scouring. The first stage in wool processing is removal of the fleece impurities, principally grease, dirt and suint, by scouring (49). The term wool grease is somewhat misleading, as *wool grease* is not a grease in the generic sense but more like a wax comprising a complex mixture of esters. The process of scouring the contaminants from raw wool can be achieved by using an aqueous medium, a solvent medium, or a combination of the two. By far the most popular type of scouring is aqueous, which is used to process 99% of the world's wool. Aqueous scouring entails washing the raw wool in aqueous solutions (neutral or slightly alkaline) of nonionic detergent or, less frequently, soap and soda, followed by rinsing in water.

The process is carried out in a series of between four and six bowls in which the wool is immersed in the wash liquor and transported through the liquor by mechanical rakes. The liquor flows countercurrent to the wool, and at the end of each bowl the wool passes through a roller squeeze. Other methods of fiber transport are used to a limited extent, eg, the wool is conveyed under jets of liquor (50) or passed around perforated drums (51), principally to limit the wool movement that leads to fiber felting. Agitation of wool in anhydrous solvents does not cause felting, and several plants (52,53) have been designed to take advantage of this property, but only a few are in operation commercially (49). The jetting of wool with high pressure solvent jets may cause the phenomenon of microentanglement, which can adversely affect subsequent processing performance (54).

The traditional conception of the contaminants on the wool has been one of fiber covered with a layer of grease in which suint and dirt (both mineral and organic) are embedded. The principal action of the detergent or soap has been seen as one involving swelling and rolling up of the grease–suint and grease–dirt mixtures and suspension of these in the water (55,56). This simple picture has been questioned by a possible fourth main component, thought to be proteinaceous in nature (57,58), which is present as a layer on the wool fiber and also as small particles dispersed with the dirt and suint in the grease. Under certain scouring conditions, eg, neutral conditions or low ratios of liquor to wool, this material does not swell readily and hinders removal of dirt. Besides affecting the efficiency of scouring if it is not completely removed, this proteinaceous layer may also influence the color of scoured wool and subsequent processing stages such as carding and shrinkproofing (58).

More recent studies (59) have shown that the proteinaceous contaminant does not form a complete layer on the fiber surface but is comprised of distinct flakes consistent with the composition of skin flakes and other epithelial debris (60). These studies identified the presence of two broad categories of contaminants on raw wool: easy-to-remove and hard-to-remove contaminants (59). The rate of removal of these contaminants in the scouring process has led to a new scouring process technology known as Siroscour, which uses the principles of multistage scouring rather than severe mechanical action to optimize the removal of contaminants during aqueous scouring. The three separate stages which comprise the heart of the Siroscour process are a modified desuinting bowl, designed to remove as much dirt as possible; a hot scouring stage that removes the easy-to-remove contaminants; and a third stage, designed to remove the hard-to-remove contaminants. In order to achieve optimum performance, separate dirt and grease contaminant recovery loops are incorporated into the Siroscour process, as well as specific bowl designs for optimum scouring of Australian wools. Several international trials have been completed, comparing different scouring systems around the world (61). The results show that in terms of entanglement, which is reflected by the final top length, Siroscour and Fleissner suction drum system perform similarly with Siroscour consistently producing wool of superior whiteness and lower residual ash content than the other systems tested.

It is important to remove as much of the wool grease from the waste scour liquors as possible, firstly because the product has some commercial value, and second, because the pollution load of the discharge is thereby reduced. Wastewater is therefore, normally centrifuged before discharge. Not all of the grease, however, can be removed by centrifuging (normally up to 40% can be removed) because certain components of the grease and dirt form stable emulsions under normal scouring conditions. Because a high concentration of dissolved solids in the liquors reduces the stability of the emulsion, it is possible under certain conditions to recover ca 85–90% of the grease originally present on the wool (62). Understanding this led to the development of the Lo-flo scouring process, in which very high concentrations of dissolved solids are produced in the scour liquors by severely restricting water input to special scouring bowls (63). The partial flocculation achieved eases the removal of contaminants by centrifuging. Grease removal is very good, but the removal

efficiencies for dirt and suint are less than normal (49), and no commercial systems have been installed.

In most situations, wastewater discharged from a scouring machine still contains large quantities of grease and water-soluble and insoluble material (both organic and inorganic). The pollution load of one modern scouring machine is equivalent to a population of 30,000 people (59). Disposal or treatment of the wastes to comply with environmental requirements is thus expensive. Probably the least expensive approach to disposal in the past has involved biological treatment by irrigation of large land areas (not less than 20 ha per scouring machine is normally required) or maturation in very large shallow lagoons (49). However, subsequent studies have shown that even this area of land is inadequate for sustainable irrigation onto normal soils cropping one to two pasture crops per year. The high levels of potassium present in scour effluent require 1000 ha of land for sustainable irrigation (64). This recognition has forced scourers to install large evaporation ponds covering many hectares. Technology has been produced to enable scourers to recover much of the potassium from the scour liquor as potassium concentrate, which can be used as a potassium supplement in fertilizer. This technology is still (ca 1997) undergoing pilot plant trials in Australia.

Physicochemical methods of wastewater treatment, eg, flocculation using inorganic or polymeric flocculants or sulfuric acid, have been investigated, as have physical techniques, eg, membrane processes and solvent extraction, but few processes have passed pilot-scale evaluation. One of the most recent processes that has proved commercially viable has been developed by CSIRO and the International Wool Secretariat and is called Sirolan CF. This is a chemical flocculation process that removes better than 75% biological oxygen demand, 80% chemical oxygen demand, and 99% suspended solids. Advantages of the process over other chemical flocculation processes are its simplicity (no large mixing or settling tanks), no inorganic coagulant required, production of a highly dewatered sludge of 60% solids, and relatively low capital and operating costs. As an add-on to the Sirolan CF process, sludge disposal technologies are being developed such as composting. This provides the scourer with an opportunity to produce a soil ameliorant or fertilizer from what was previously a waste material. Research in the 1990s is aimed at introducing scour waste integrated management systems into the industry that reduce consumption of materials such as water, chemicals, and energy to an absolute minimum and that recycle and reuse any waste materials as usable by-products. One such package, known as SIROSWIMS, is designed to reduce water consumption to almost zero with the production of a potassium-rich concentrate for fertilizer use and a sludge that can be composted or burned (65).

One physical method that has attracted some commercial interest is evaporation; several evaporative plants were installed in Japan in the early 1970s, nearly all followed by incinerators for the sludge produced (60). They are, however, expensive in both capital and operating costs. The most recent evaporation systems use a process known as vapor recompression, which has the claimed advantage of much lower operating costs than the earlier evaporative processes used in the wool industry. Capital costs of these processes are still high.

Another approach to effluent treatment is to reduce the effluent load and volume by modifying the scouring process. This is the basis of the Lo-flo process mentioned and also of the Wool Research Organization of New Zealand (WRONZ) rationalized scouring system (66–68). The SIROSWIMS process referred to also segments the scour effluent into different streams suitable for particular treatments; for example, the suint flowdown is concentrated to the potassium-rich fertilizer supplement; the scour effluent strong flow is treated in the Sirolan CF process to produce a clarified effluent and a sludge suitable for composting or burning; and the rinse discharge can be filtered and then recycled (65).

Carbonizing. Carbonizing is a process used to remove excessive contents of cellulosic impurities, eg, burrs and vegetable matter, from wool. It is carried out on loose wool (65), rags (69), and fabric (70). With loose wool and fabric, the wool is treated with aqueous sulfuric acid and then baked. The cellulosic matter is rapidly destroyed by the hot concentrated acid, by which it is converted into friable hydrocellulose, whereas wool is scarcely affected by the treatment. For rags, hydrogen chloride gas is preferred as the mineral acid because it has less effect on the colors. After acid treatment, the carbonized vegetable matter is crushed to facilitate removal. The wool is then normally neutralized in alkali, although some mills omit this stage to facilitate subsequent dyeing with equalizing acid dyes.

Most industrial carbonizing is done on loose wool, and the technology has changed little over the last 40 years. Australian carbonizers in the 1990s use surface-active agents to reduce fiber damage in the carbonizing process (71,72). The improved results depend on the use of high acid concentrations and rapid throughput (71). Factors important to the role of the surfactants are thought to be the increased efficiency of acid removal in squeezing (73) and the improved spreading of the acid over the fibers in the presence of the surfactant (74).

Attempts have been made to introduce processes and equipment for rapid acidizing, drying–baking, and crushing (75) and for carbonizing wool in sliver (rope) form (76), but neither of these processes has been adopted commercially.

Developments in the carbonizing of fabric (77) have been aimed at reducing the volume of acid used (78,79), in order to reduce energy in drying and save neutralization costs, and at the use of generally available textile processing equipment, eg, the padmangle, for application of the acid in a pad-dry-bake process (80).

Processing/Combing

A wool combing plant typically comprises the following processes:

Process	Product
scouring	scoured wool
fiber lubrication	
carding	card sliver, burr wastes, card droppings
first preparer gilling	prepared sliver
second preparer gilling	combed sliver, noil
third preparer gilling	
combing	top
first finisher gilling	balls of top
second finisher gilling	bumps of top
packaging	broken top for woolen processing

For very dusty wools or wools with high vegetable matter (VM) content, machines with a beating action may be used to remove some of these contaminants, either from greasy wool or scoured wool. Unfortunately, nonwool contaminants may also be present. Bale twine (polypropylene) and bale pack material (usually polyethylene) and articles of clothing are examples. These become fiber-like in processing and may cause costly problems in fabric. The common way to remove some of these is to station pickers at the scour output, but their efficiency is low. For high quality markets the careful selection of sources of raw wool may be necessary. The successful development and introduction of the SIROCLEAR system (81,82) for the detection and removal of dark fiber in yarns is appropriate to very low contaminant levels. Yet SIROCLEAR does not remove the necessity for quality control at the early stages.

Distribution of the scoured wool to the cards then occurs. This is often done using a combination of pneumatic and mechanical transport systems, usually with the provision of accumulator bins that hold a few hour's production. Modern cards are usually fed with tower feeds, which automatically control the height of the wool column.

Carding is the process of individualizing the entangled fibers in scoured wool and reassembling them into a "sliver" weighing about 25 kg km. At the same time, up to 90% of the VM is removed. Most fiber breakage in early-stage processing occurs during carding and the level and type of lubricant applied prior to carding can change the average fiber length by several millimeters (83). The average fiber length, or "Hauteur," of the top is an important quality parameter to the spinner. A change of several millimeters in the final result is associated with a change in the value of the top by 1% or so. The Hauteur is reduced and the combing waste, or "noil," increases as the card production rate increases, limiting card production rates. Modern cards have production rates of about 120 kg h for 21 μm wool, and compared to cotton carding they have a low production per unit cost. CSIRO research has led to

significant productivity gains in carding, and in 1995 a new series of cards was released by Thibeaudeau, who collaborated with CSIRO and IWS in the development of cards having production rates in excess of 200 kg h (84,85).

The efficiency of the vegetable matter (burr and seeds) removal mechanism depends on the careful maintenance of settings and speeds and the level of drying of the scoured wool (86). The burr wastes contain wool fiber. Burr, together with fiber which has dropped beneath the card, may be carbonized and used in the woolen system.

The next three stages after carding, those of preparing for combing, use a process of drawing called gilling. Six to eight slivers are submitted to two consecutive sets of rollers. The output, or delivery set, of the rollers has a surface speed of between six and eight times that of the input, or feed set. Rows of pins attached to bars, located between the rollers, are inserted into the slivers from above and below. These travel forward at a speed marginally greater than that of the feed rollers. The delivery rollers, whose speed is much greater than that of the pinned bars, draw the fibers through the pins, thus improving the fiber alignment along the axis of the emerging sliver.

Because of the multiple sliver feed to the gills, the three preparing stages achieve a further degree of blending. This is valuable because the entire consignment of top is never blended as one batch as in woolen processing. There is no sophisticated bale-opening equipment prior to carding that samples multiple bales at one time, as in cotton processing.

A major effect of the use of three gilling stages is the reduction of the waste or noil produced in combing. While more stages would further reduce noil, the cost to the mill would outweigh the gains.

At least one of the preparer gills will incorporate autoleveling. After the thickness of the feed slivers is sensed, adjustments to the speed of the delivery rollers are automatically made to correct deviations from the desired average. The aim is to achieve as high a degree of uniformity as possible in the slivers at the entry to the comb.

Japanese customers in particular may prefer tops that have been produced with a backwashing stage. This is a gentle aqueous sliver scour incorporating several bowls and a dryer following the first preparer stage. Its purpose is to remove carding lubricant and additional dirt and protein contaminant. It may increase the Hauteur and is an effective way to add moisture to the slivers for combing. It also incurs a significant added expense.

Modern gilling is highly productive, largely as a result of improved drive mechanisms for the pinned bars. A modern machine can deliver at 400 kg h.

Process 7, that of combing itself, employs a very sophisticated piece of machinery, shown schematically in Figure 6. Gripped by jaws or nippers, the leading fringe of fibers of the prepared slivers is combed by a cylinder whose surface is clothed with pins or teeth. This is the circular combing phase, and the short fiber, entanglements, and VM gathered by the circular comb constitute the bulk of the waste or noil produced in combing. The top or front comb is then inserted into the circularly combed fringe, and the tip of the fringe is gripped by the detaching rollers. These pull the beard of fibers through the top comb and lay it like a shingle on the previously withdrawn beards to recreate a sliver. In this detaching phase, the trailing ends and the fibers are combed and the small amount of fiber and waste which accumulates in the top comb is also added to the noil. The entire length of the fibers is thus combed, either by the circular or top combs.

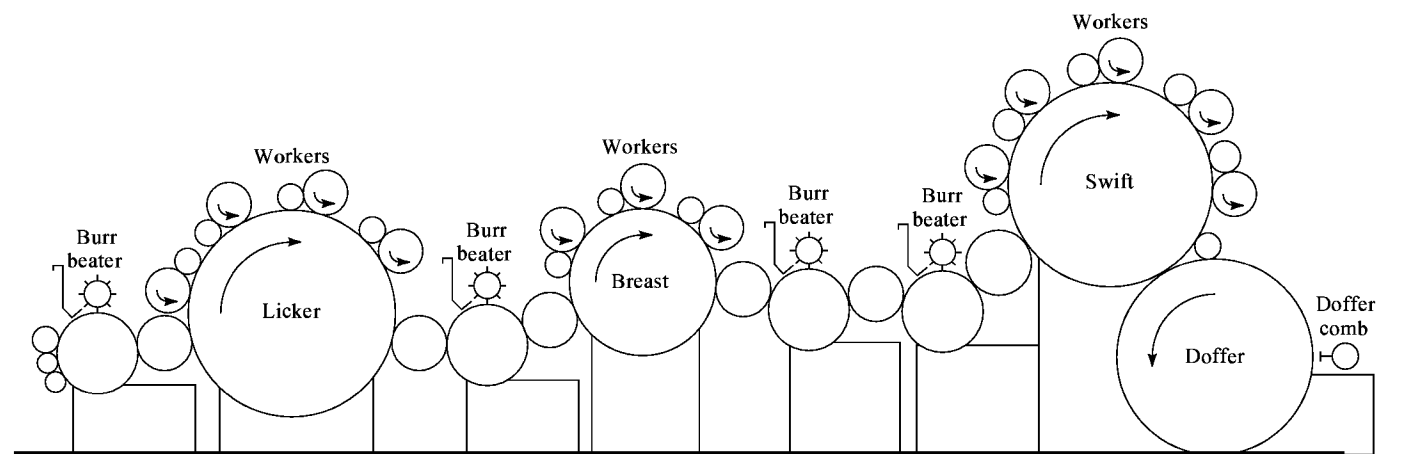


Fig. 6. Schematic of the wool combing process.

The comb removes vegetable matter (VM) with an efficiency of about 99%. It also removes very short fiber less than about 20 mm in length and tight entanglements or neps, again with very high efficiency. Vegetable matter and neps can cause problems in later processing and faults in knitted or woven fabric that are costly to remove. Short fiber causes fly problems in later processes, ie, accumulations of loose fiber which can lead to objectionable faults in yarns and fabrics.

Combing is thus a critical process in producing a quality top. It is also a costly process, and a modern comb will only produce at a maximum rate of about 30 kg h for 21 μm wool.

Processes 8 and 9 again employ gills. The purpose is to shuffle the fibers in the combed sliver to produce a random positional distribution, and thus the most even sliver possible. The second finisher gilling will also incorporate either a ball winding head, or a press that condenses top that has been delivered into a can. This latter product is called bump top. Both ball and bump facilities are usually necessary to satisfy customers' differing needs. Bumps or balls will then be further pressed into bales for shipment. Combed wool destined for processing on the woolen system may be sucked in short lengths from the comb to form broken top.

The transport of slivers from the card through the various processing stages to the final finisher gill is achieved either by using wheeled cans up to 1.5 m in diameter or by winding balls. Automatic systems for materials handling are at the prototype stage, but will be part of the combing plant of the future. In this way a topmaking plant will approach a continuous process rather than a batch process and labor costs will reduce as a result.

When processing wools with a diameter of 21–23 μm, a scouring line would typically produce 1000 kg h. To handle this output requires eight cards, and two lines of at least three preparer gills. Depending on materials handling and autoleveling requirements, 30 to 40 combs would be needed, and another two lines of two finisher gills together with bumping, balling, and baling equipment are required. A backwash may also be needed. Production of such a mill would be about 5000 t annually.

Some combing plants also produce shrink-resistant treated wool-top. The current process incorporates chlorination, and the application of a shrinkproofing polymer. Chlorination is likely to be replaced soon by more environmentally friendly treatments.

Quality control requires a testing laboratory where certain characteristics are measured. These include Hauteur and fiber length distribution, diameter, diameter distribution using Sirolan–LASERSCAN, solvent-extractable level or fatty matter content, nep, VM, dark fiber content, moisture content or regain, color, and sliver linear density and its uniformity. In the future, automatic measurements of many of these parameters will occur on-line at several stages in the mill, and not just at top stage. The collected data will be part of increasingly computerized mill management systems.

Worsted Spinning

The worsted spinner converts the top of about 25 kg km linear density into roving of between 200 and 2000 g km, depending on the linear density of the

yarn to be spun. The sequence of processes incorporates four or five drawing stages. The first three or four usually employ gills, whereas the last is commonly carried out on a roving frame in which the attenuated strands are rubbed to impart cohesion.

For finer yarns, where freedom from nep and VM is critical, there are initially two stages of drawing. These are followed by a recombining operation, then by two further passages of drawing to reform the top before the previously described sequence begins. Blending of synthetic fibers with fine wool requires a similar preparation also incorporating a recombining operation.

On the spinning frame, the rovings are again attenuated in a drawing operation, usually with a draft ratio of about 20:1, twist is inserted, and at the same time the yarn is wound onto bobbins. The yarn on the bobbins is usually steam-set in an autoclave to reduce its twist liveliness, then wound into packages or cones. Detection and removal of thick or thin spots occurs during winding. SIROCLEAR can also detect and allow the removal of dark fiber or VM contamination. Following twisting into a two-fold yarn and further setting, the yarns are ready for knitting or weaving.

Spinning and twisting are very costly operations because of their low productivity: typically 40 spindles produce one kg h of yarn. Spinning technology is becoming highly sophisticated, automated, and increasingly under computer monitoring and control. Quality is paramount in producing a competitive yarn in a highly competitive commodity market. Regularity, strength, extensibility, and freedom from imperfections and contaminants are the important parameters.

Woolen Spinning

In woolen spinning there are no highly efficient mechanical methods to remove VM. Generally, very clean scoured wool, combed wools, or carbonized wool must be used as inputs, or fabrics must be carbonized.

The web of fibers at the card delivery is split longitudinally into about 120 ends, rubbed to impart cohesion, and wound into cylindrical packages called slubbing. Twisting of these slubbings occurs on the woolen spinning frame with very little draft, so as to produce woolen singles yarn. This yarn is weavable without any further processing other than a winding and clearing operation similar to that used in the worsted sequence.

Woolen spinning is thus a very short sequence compared to worsted processing. Woolen yarns are economically uncompetitive with worsted yarns even at relatively coarse yarn counts, a result of the yarns being virtually made on the carding machine, the expense of the much larger card used, the low delivery speed, and the very low card production in spinning fine yarns, resulting from the low delivery speed.

Woolen yarns cannot be spun as fine as worsted yarns, even when using the same fiber diameter. The fabric weights are greater and they have a harsher feel. For these reasons woolen products have not been able to follow the modern trend to smoother lightweight clothing as easily as their worsted counterparts.

The woolen processor is relatively close to the consumer market. The essence of its success is the identification of lucrative markets in apparel or home finishings and the appropriate choice of the most cost-competitive blend fiber input. Technical experience, skillful design, and effective marketing are mandatory.

Dyeing

Wool is dyed from aqueous solutions. The majority of dyes used on wool are sodium salts of aromatic anions. Water solubility is usually provided by sulfonic acid groups, but in a few cases carboxyl or hydrophilic, nonionic substituents are used.

Wool has a complex chemical structure, composed mainly of a large number of different proteins (87). It is amphoteric in character because of the presence of basic amino and acidic carboxyl groups in the side chains of some of the component amino acids. In an aqueous acidic dyebath, protonation of the amino and carboxyl groups results in a net positive charge on the fiber.

Early workers explained the mechanism of wool dyeing in terms of electrostatic interactions between the positively charged amino groups on the wool and anionic dye molecules (88). The two most popular theories were the Gilbert-Rideal theory and the Donnan theory (89,90). Later studies, however, showed that the affinity and wetfastness properties of dyes on wool are largely determined by other types of interactions, namely van der Waals forces and interactions between hydrophobic regions in the fiber and hydrophobic parts of the dye molecules (91,92). Dyes differ markedly in their hydrophilic–hydrophobic character (92–94). Ionic forces are of greatest importance for dyes of relatively low molecular weight. With increasing molecular weight, van der Waals and hydrophobic interactions become increasingly important; this is reflected in higher affinity for the fiber and better wetfastness properties. However, ionic interactions are important for all wool dyes in determining the rate at which the dye is taken up by the fiber from the dye liquor.

The dyeing rate can be controlled by varying the amount of acid added to the dyebath, because this controls the size of the net positive charge on the wool, produced by protonation of amino groups. Early workers studying the uptake of dyes by wool treated the fiber as a cylinder of uniform composition. These studies were mainly concerned with the thermodynamics of the dyeing process, in particular the situation applying when equilibrium had been attained. They provided little information, however, on the mechanism of the dyeing process itself. Recent work has clarified the relationship between fiber structure and the mechanism of wool dyeing (95,96).

Wool fibers have a very complex morphological structure. They can be considered as biological composite materials, in which the various regions are both chemically and physically different (87). Fine wool fibers contain two types of cell: those of the internal cortex and those of the external cuticle. Cortical cells, which constitute around 90% of the fiber, are spindle-shaped and are arranged in an overlapping pattern, parallel to the fiber axis. They are separated and held together by a continuous network, the cell membrane complex. The external cuticle cells are rectangular in shape and overlap, rather like tiles on a roof. Dyes enter wool fibers at the junctions where the cuticle cells overlap and in the early stages of the dyeing process they diffuse into the interior along the network of the cell membrane complex. Later in the cycle, the dyes transfer, progressively, from the cell membrane complex into the cortical cells. The affinity of the dyes for wool is largely the result of their interaction with hydrophobic proteins located within the cortical cells.

The dyes used on wool can be divided into the following groups: acid dyes, chrome dyes, premetalized dyes and reactive dyes (88,89,92–94,97–99). Strictly speaking, all types of wool dyestuffs can be described as acid dyes, but in practice this term is confined to leveling acid dyes, half-milling dyes, milling dyes, and supermilling dyes (94,97). This subclassification of acid dyes arises out of the methods used for their application and their fastness properties on wool.

Leveling acid dyes are the simplest type of dyes used on wool. They are of relatively low molecular weights (300–600) and have a comparatively low affinity for the fiber, because they interact largely by ionic attraction. They are applied at low (ca 3) pH values, obtained with sulfuric acid. Under these conditions, the initial uptake of dye is usually uneven because the exhaustion rate is very rapid, but the dyes have good migration properties and leveling occurs during the dyeing cycle to produce a very even result. Leveling of the dyes is assisted by the presence of sodium sulfate in the dye bath, because the anionic sulfate ions compete with dye anions for the positively charged amino groups in the fiber. In general, these dyes are not particularly resistant to wet treatments, but are used when level dyeing is critically important.

Milling acid dyes are of higher molecular weight (600–900) and are more hydrophobic than leveling acid colors, which gives them a higher affinity for wool. The higher molecular weight and lower dependence on ionic attraction means that they diffuse more slowly and, therefore have higher wetfastness than leveling acid dyes. Their migration properties are, however, not as good as acid dyes, and in order to obtain level dyeings it is important to ensure that exhaustion is uniform and not too rapid. This is achieved by setting the dyebath at a higher pH (ca pH 4.5–5.5) with a weak acid, such as acetic acid, and also by the addition of a surfactant-type leveling agent. Half-milling dyes fall between leveling acid and milling dyes in molecular weight (500–600), migration properties, and wetfastness. They are applied at a pH in the range pH 4–4.5.

Supermilling acid dyes are similar in molecular weight to milling dyes, but contain long alkyl groups that make them more hydrophobic in character. This gives them a high affinity for wool that is virtually independent of ionic interactions with the fiber; consequently, they have good wetfastness properties. They show good exhaustion under almost neutral (pH 5.5–6.5) dyeing conditions, but their high affinity means that they have poor migration behavior. Auxiliary products are often used to assist leveling of these dyes.

Chrome dyes are acid dyes (mol wt 300–600) that contain groups capable of forming complexes by reaction with a metal salt, usually sodium or potassium dichromate (94,97). The chrome dye complex has lower solubility, and hence better wetfastness, than the parent dyestuff. Reaction between the

dye molecule and chromium salt can be carried out before, during or after application of the dye to the wool. Modern practice is to carry out the chroming step after dyeing. Chrome dyes are relatively cheap, have good migration and level dyeing behavior and excellent wetfastness properties. Their popularity, ca 1997, except for black and navy blue shades, has declined in recent years because of the need for prolonged dyeing cycles, fiber damage associated with oxidation of the fiber in the chroming step, and environmental concerns about the use of chromium salts.

In many applications, chrome dyes have been replaced by metal-complex dyes, which have a very high affinity for wool (94,98). In these dyes, the metal complex is preformed by reaction of one metal atom with either one (1:1 metal-complex dyes) or two (1:2 metal-complex dyes) molecules of dyes that contain groups capable of coordinating with chromium, or occasionally cobalt atoms. In general, metal-complex dyes produce duller shades than acid or milling dyes. The 1:1 metal-complex dyes are applied to wool from strongly acidic dyebaths at around pH 2. They are almost all monosulfonates (mol wt 400–500) and have good leveling behavior and wetfastness properties. Some degradation of the fiber can be caused by the low dyeing pH. The earliest type of 1:2 metal-complex dyes were unsulfonated; solubility was provided by nonionic polar groups, such as sulfonamide. More recently, monosulfonated, disulfonated, and some carboxyl-containing types have become available. These dyes range in molecular weight from 700 to around 1000. They are applied at pH values ranging from pH 4.5 to 6.5, depending on the degree of sulfonation and molecular size of the dye.

Reactive dyes are becoming increasingly important for dyeing wool (93,99). They have molecular weights in the range 500–900 and usually contain two or three sulfonic acid groups. These dyes also contain groups that react covalently with wool, which gives them outstanding wetfastness properties. They are characterized by bright colors and moderate migration and leveling properties, provided careful attention is given to ensuring that the exhaustion rate is not too rapid. This is achieved by careful control of the dyebath pH and by using special amphoteric leveling agents. Their ability to react with the fiber means that an unlevel strike of these dyes is very difficult to rectify. Reactive dyes are relatively expensive and their most important application is on wools that have been treated to withstand shrinkage in machine washing.

Wool is usually dyed as loose fiber, or in sliver, yarn, fabric, or garment form by batch methods (88,100). In the case of fiber, sliver or yarn, the process involves pumping dye solution through the stationary substrate; whereas in fabric dyeing the material may be transported through the dye liquor or, alternatively, both liquor and fabric may be moved through the machine. Simultaneous circulation of liquor and wool substrate is also used in garment dyeing techniques. In a typical operation, the liquor is heated from around 40°C to a temperature of 98–100°C, where it is held for a period that can range from around 30 min to over 2 h, the actual time depending on the type of dye and depth of shade. Procedures have also been developed for continuously dyeing wool in loose or sliver form and wool fabric can be dyed by a pad-batch method (93). However, because of the need for large volumes of one color for economic operation, these techniques have found only limited use.

When wool is dyed for prolonged times at the boil, the fiber can be damaged as a result of hydrolysis of the component proteins. Recently, the Sirolan-LTD Process for dyeing all types of wool substrate at temperatures of 85–90°C has been released (101). The method employs a special chemical that increases the rate of exhaustion and diffusion of dye into the fiber, thus enabling satisfactory dyeings to be obtained at temperatures below the boil. Wool dyed by this method suffers less degradation than wool dyed for a similar time at the boil. This leads to benefits in the processing performance of the dyed fiber and in improved quality of end products (102). Permanent setting of wool fibers during dyeing has been identified as a major factor that also contributes to deterioration of the quality of dyed wool (102). This effect is separate from hydrolytic degradation. Setting occurs via a mechanism involving thiol–disulfide interchange. Methods of inhibiting the permanent setting of wool fibers during dyeing have been developed (103). These involve addition to the dyebath of auxiliaries (antisetting agents such as Basolan AS) that block thiol groups and thus prevent thiol–disulfide interchange. Improvements in the properties of a range of products made from wool are obtained by reducing the amount of permanent set that occurs during dyeing.

Printing

Compared with cotton fabric, 40–50% of which is printed, the fraction of the wool clip used for printing (~1%) is minute. The principal reason is that cotton (qv) is usually pigment printed, which involves a straightforward print–dry–thermal curing process. Pigment printing of wool is insignificant owing to the detrimental effects of the pigment binder on the handle, softness and drape of wool fabrics. Printing of textiles using dye pastes allows these qualities to be maintained, but requires a far more complex processing sequence involving several stages. The usual route is fabric preparation–print–dry–steam at 100°C (to fix dye to fabric)–wash off (to remove thickener and unfixed dye)–dry (104). However, an estimated 20 × 10⁶ m of high quality printed wool fabric is produced in this way per annum worldwide, the center of excellence being in northern Italy. Reference 104 provides reviews of wool printing.

Fabric preparation is often considered to be the most important stage to obtain good color yields, levelness, and brightness on wool fabric (104). This is done almost exclusively by an oxidative chlorination process, the most popular commercial methods using either a batch treatment with dichloroisocyanuric acid (DCCA), or a continuous fabric treatment with gaseous chlorine, called the Kroy process.

Chlorination has a profound effect on wool, especially near the fiber surface. Cystine residues are oxidized to cysteic acid, and accessible peptide bonds are cleaved at tyrosine residues, forming carboxyl groups, thus increasing the number of anionic groups. Chlorination also reduces the degree of cross-linking, which increases rates of diffusion, especially in regions of the fiber where cystine residue concentrations are high. Soluble protein is released from chlorinated wool fibers, which become readily wettable, in contrast to their hydrophobic nature before treatment.

Chlorination can have several adverse effects, especially at higher (4% owf DCCA) levels, where problems in obtaining level treatments can arise. Yellowing, especially after steaming, is also of concern, and these difficulties led to the development of the continuous Kroy fabric chlorinator. However, in all chlorination processes, soluble protein and free oxidation products must be removed from the wool by neutralization with a reducing agent such as sodium bisulfite or sodium dithionite before drying. This effluent from chlorination of wool contains absorbable organohalogen compounds (AOX), which is regarded as undesirable for release into the environment, particularly in Europe. Several AOX-free methods for preparing wool for printing are under investigation.

Plasma (105) and corona discharge (106) treatments have been trialed as possible prepare-for-print methods for wool, as they oxidize wool fiber surfaces and have the advantage of being dry processes. Although the wettability of fabrics prepared using these methods is high, the color yields of prints are significantly less than prints on chlorinated fabric, and the fabric handle is inferior. An alternative approach is the Siroflash process, which involves continuous ultraviolet (uv) irradiation of dry wool fabric with a commercial uv source used for curing polymer films, followed by conventional oxidative bleaching with hydrogen peroxide (107). This process is surface-specific, giving a white ground fabric with good handle properties, and similar color yields to DCCA (4% owf).

Most wool is printed directly onto the prepared ground fabric. Because of the shorter run lengths normally processed for wool fabrics, automatic flat-screen printing remains the most popular method over the rotary screen process used for most cotton printing. The dyes most commonly used for wool printing are acid milling, metal complex, and reactives. A simple dye paste is prepared using a suitable thickener, urea to swell wool fibers during steaming, and dye. Various additives may be included in the paste recipe, such as antifoams, glycerol to prevent screen blocking, an acid donor to ensure printing at low pH, wetting agents and dye solvents, such as thiodiglycol (104).

During the steaming stage large aggregates of dye molecules break down into smaller entities which can then penetrate the swollen wool fibers, and fixation occurs. Typical steaming times are 10–15 min for reactive dyes, and 30 min for acid milling and metal complex dyes. Saturated steam is essential for optimum penetration and fixation of dyes, especially for dyes of high molecular weight such as copper phthalocyanines. Both continuous and batch steamers are used commercially. Chlorinated fabrics are highly prone to yellowing during steaming, which can be detrimental to printed pale shades.

The aim of washing off is to remove the thickener and any residual chemicals and unfixed dye from the fabric, without causing staining of any pale or unprinted ground areas. Prints using metal complex and acid milling dyes are usually washed off in water at 30–40°C, whereas prints using reactive dyes require higher (60–80°C) temperatures with addition of ammonia (104). Both open-width ranges and winches are used commercially to wash off wool prints.

Discharge printing of wool remains popular despite its technical difficulties (104). In this style of printing, a predyed fabric is printed with a reducing agent that destroys the ground shade. However, included in the print paste is a dye which is resistant to the discharge agent. The result is bleaching of the

ground color and replacement with the illuminating color in printed areas. It is often impossible to reproduce the designs and effects of discharge printing with direct printing. The reducing agents used in discharge printing are zinc, calcium or sodium formaldehyde sulfoxylates or thiourea dioxide. The technique is limited by the range of dyes that are resistant to the reducing agent.

Other techniques, such as resist printing, cold batch printing, and transfer printing, have previously been applied to wool with some success, but these occupy only a small share of the current wool printing market (104).

Bleaching

Reduction of the natural cream-yellow color of wool is sometimes necessary for improving the brightness of dyed shades, particularly pastels, and for improving the ground color of wool fabrics before printing and after chlorination. Commercial wool bleaching is carried out using oxidation, reduction, or combined oxidation–reduction processes, and oxidation in the dyebath is also possible (108). In general oxidative bleaching with hydrogen peroxide gives superior whiteness over reductive methods. Reference 108 provides reviews on wool bleaching.

For most bleaching applications a batch treatment with hydrogen peroxide is used, and an activator (or catalyst) is normally required to increase the rate of bleaching. The most common activator is alkali, and wool bleaching at pH 8–9 for 1 h at 60°C with a stabilized 0.75% peroxide solution is a typical process (109). Under alkaline conditions it is generally accepted that the active bleaching species is the perhydroxy anion OOH[−], the formation of which is encouraged at higher pH (110). Peroxide bleaching of wool under mild acid conditions (pH 5–6) can also be carried out using a peracid activator such as Prestogen W (BASF) or citric acid (111). As wool sustains some damage in the presence of alkali, this method can be useful when it is necessary to bleach more delicate fabrics.

With hydrogen peroxide, a stabilizer is required to reduce its decomposition to water and oxygen. This reaction is catalyzed by transition metal ions, particularly iron, copper, and manganese, and therefore most stabilizers act as sequestering agents for these metal ions. The most common stabilizers for alkaline wool bleaching are phosphates, particularly tetrasodium pyrophosphate, and a typical dosage for this is 6 g/L of bleaching liquor. However, recent concerns over phosphates (ca 1997) in effluents have led to the development of alternative stabilizers based on silicates (112). IWS have examined an alkaline peroxide bleaching process using the silicate-based Stabicol WDI (Allied Colloids) (113).

Heavily pigmented wools such as karakul require a more stringent approach, known as mordant bleaching, in which a metal salt is first applied. The metal cations are preferentially absorbed by the melanin pigment, where they subsequently decompose hydrogen peroxide to produce highly aggressive hydroxyl free radicals, which then attack and bleach the melanin (114).

For reductive bleaching of wool the two most popular chemicals are stabilized sodium dithionite (sodium hydrosulfite, CI Reducing Agent 1) and thiourea dioxide (CI Reducing agent 11). Most reductive bleaching of wool is carried out using stabilized dithionite (2–5 g/L) at pH 5.5–6 and 45–65°C for 1 h. Thiourea dioxide is more expensive than sodium dithionite, but is an effective bleach when applied at the rate of 1–3 g/L at 80°C at pH 7 for an hour.

Whiter fabrics are produced if an oxidative bleaching is followed by a reductive one; this is often referred to as full bleaching. One such process involves a hydrogen peroxide oxidative bleaching, followed by addition of thiourea to the liquor to generate thiourea dioxide for a reductive bleach *in situ* (115).

Yellowing of wool can occur during dyeing, depending on pH, temperature and time, and chlorinated wools are especially sensitive. Bleaching agents that can be added to the dyebath have been developed based on sodium bisulfite and hydroxylamine sulfate (108). Addition of hydrogen peroxide to the dyebath after exhaustion can also be effective.

Whitening

Wool fabrics that have been bleached or treated with fluorescent whitening agents (FWAs) yellow rapidly in sunlight, especially when wet. This is a major problem when bright pastel shades are required, and despite a significant amount of research into the chemistry of photoyellowing and photobleaching processes (116), progress has been limited.

The use of some water-soluble benzotriazoles, a well-known class of uv absorber, can be effective against both photoyellowing and phototendering. These compounds rely on intramolecular hydrogen bonding and proton transfer to be most effective, and in early studies intermolecular bonding to wool fibers was found to be highly detrimental to their performance. Optimization of the photochemical properties of soluble benzotriazoles has resulted in a suitable formulation, Cibafast W, for use on bleached and normal wool (117). However, uv absorbers cannot be used on FWA-treated wool, since they absorb the wavelengths necessary to cause fluorescence.

Recent work has suggested that when wet FWA-treated wool is exposed to sunlight, hydrogen peroxide is formed (118). As doping, FWA-treated wool with hydrogen peroxide increases the rate of photoyellowing significantly, blocking the formation of peroxide could improve the photostability of whitened wools.

An alternative approach is to physically separate the FWA from the wool fiber by incorporating the whitener into a suitable surface polymer treatment for wool fabric (119). The photostability of these fabrics is significantly better than conventional FWA treatments, being similar to bleached wool.

Insectproofing

Wool, as a keratin, is a highly cross-linked, insoluble proteinaceous fiber, and few animals have developed the specialized digestive systems that allow them to derive nutrition from the potential protein resource. In nature, these few keratin-digesting animals, principally the larvae of clothes moths and carpet beetles, perform a useful function in scavenging the keratinous parts of dead animals and animal debris (fur, skin, beak, claw, feathers) that are inaccessible to other animals. It is only when these keratin-digesting animals attack processed wool goods that they are classified as pests. Very often they enter domestic or industrial buildings from natural habitats such as birds' nests.

The principal insects that attack wool are the common clothes moth, *Tineola bisselliella*, the case-bearing clothes moths, *Tinea metonella*, *Tinea dubiella*, *Tinea transluens*, and *Tinea pellionella*, *Monopis crocipitella*, the brown house moth, *Hofmannophila pseudospretalla*, and the variegated carpet beetle, *Anthrenus verbasci*. The taxonomy of the *Tineid* species has recently been comprehensively reviewed (120). Studies in Australia have shown that the native *Tineids* are rarely involved in domestic infestations, and are predominantly found in nests of native bird species. The major pest species that are found in dwellings in Australia are usually the introduced species, such as *Tineola bisselliella*, *Tinea transluens*, and this *Tinea* species is often associated with the nests of the introduced urban bird species such as sparrows and swallows. *Tineola* is reported to live mainly in human environments (121).

In devising strategies to prevent insect attack on processed woolens, it is appropriate to aim only to protect processed wool goods and to control the principal domestic pest species. This is usually achieved by incorporation of an insecticide into those goods required to be protected, usually from the dyebath while the articles are being dyed, and methods for protection of wool goods have been reviewed (122).

The use of the dyebath as an application medium is attractive, not only because it avoids the needs for an additional wet processing step but also because the pesticide has a high probability of being taken inside the highly swollen fiber. This ensures that the pesticide will be relatively fast on the wool, as diffusion, especially from the dry fiber, will be slow. As an example, permethrin, the main active agent currently used for insect-resist treatments of wool goods, has a half-life of only 19 d in storage at 60°C when sprayed onto wool, but its lifetime exceeds the usual life of wool goods when applied from boiling dyebath (123). With the pesticide located inside the fiber, there is little contact action against the target pests, and the active agent is most efficiently released only in the insect gut when the fiber is completely degraded.

Unfortunately, there is a significant disadvantage resulting from application of insect-resist agents from dyebaths: it is impossible to ensure 100% exhaustion (transfer of pesticide from dyebath to fiber) and as a result, there is inevitably some environmental contamination. The extent of concern with this release of insect-resist agent depends on the spectrum of activity of the agent. If it is a broad-spectrum insecticide, especially one with reasonable persistence and lipophilic character, it is liable to be reasonably toxic to aquatic insects and invertebrates, especially in certain environmental locations where

large amounts of wool are treated with insect-resist agent. It is for this reason that restrictions on the use of pyrethroid-based insect-resist agents are being imposed, and U.K. authorities have set an allowable environmental concentration for permethrin of 10 ng L (124–126).

This requirement for environmental safety adds an almost impossible requirement to those already demanded of an insect-resist agent. Ideally, an insect-resist agent for permanent protection of wool goods should be highly toxic to both moth and beetle species, allowing low levels of application; it must be stable to the application conditions (high temperature in neutral to acid aqueous solutions), resistant to washing and light, safe to consumers, resistant to leaching from the fiber by perspiration and saliva, and effective for prolonged periods on the textile, ie, at least 15 years for wool carpets. It must be amenable to efficient application to wool under a wide variety of dyebath conditions, in the presence of dyes, dyeing assistants, and surfactants, some of which may decrease the efficiency of application by increasing the water solubility of the insect-resist agent.

A wide variety of broad-spectrum agricultural insecticides have been investigated for their suitability for insectproofing. Other chemicals investigated have included juvenile hormones, compounds, anionic surfactants, organotin compounds, eg, triphenyltin chloride and acetate, thioureas, and 2-thienylethylene (127,128). Some of these have shown activity against wool-eating insects, but all, with the exception of certain synthetic pyrethroids, have disadvantages that have precluded their commercial dyebath use. Certain quaternary ammonium detergents have found specialized use for protection of wool insulation, where they provide a combined rot-resist and insect-resist action. Because of their ionic character, they are not lost from the wool by volatilization when applied by spraying (123).

One novel approach involved modification of certain organophosphorus insecticides that showed activity against wool-eating pests but suffered from problems of volatility and poor affinity for wool. A group that reacted covalently with the wool fiber was bonded through an ester linkage to the insecticidal molecule. This covalent bonding improved the fastness of the treatments, and the ester linkage allowed cleavage of the active insecticide in the digestive tract of the wool-eating pests. This poisoned the insect during digestion of the treated wool. Commercial products based on the principle have not been developed (129).

Permethrin-based insecticides have been the principal formulations used for insect-resist treatments for wool since the late 1970s, and their cost has progressively declined. Prior to the introduction of synthetic pyrethroids, insect-resist agents were mainly based on specialty materials such as polychloro chloromethyl sulfonamido diphenyl ester (PSCDs) (Eulan U33, Eulan WA new) but their production has been discontinued because of a combination of environmental toxicity (126) and cost. Production of insect-resist agents based on the synthetic pyrethroid cyfluthrin (Eulan SP) has also been discontinued. One specialty insect-resist agent based on sulcofuron (Mitin FF) released in 1939 (130) retains a small market share for specific applications where cost is not important, and where high resistance to wet conditions is required. It is also useful on wool insulation where it also has low volatility and provides resistance to rotting. Methods for analysis of these agents on wool have been developed for quality control purposes (131,132).

It is likely that broad-spectrum pesticides will continue to have a cost advantage over specialty insect resist agents because the broader sales base will likely help to defray development and registration costs, and competition may also force prices down. High development costs will make it difficult to develop future specialty insect-resist agents, as the market is too small to support the kind of toxicity and safety testing required to demonstrate the safety of modern, biologically active agents. There are two general approaches toward dyebath application of agricultural insecticides that are being pursued: (1) the pyrethroids considered environmentally safe, such as cycloprothrin (133). Unfortunately, the safer pyrethroids also have lower activity than the mainstream pyrethroids, so that they need to be applied to the wool at higher concentrations and therefore at higher cost. (2) There are some prospects for development of species-specific pesticides such as nitromethylene (high beetle activity) (134). These are being developed, ca 1997, to fill specialized needs in general agriculture, but for insect-resist protection of wool they need to be mixed with other active agents to provide protection against the full range of keratin-digesting pests.

Because of the difficulty of developing new insect-resist agents for dyebath use, there is increasing dependence on permethrin, almost the sole agent in current (ca 1997) use. This not only increases the volume of environmentally unacceptable discharges, but also increases the probability that resistant species will emerge. Resistance to dieldrin developed rapidly in Australia at a time when it was widely in use (135,136). Resistance to pyrethroids is common among other agricultural pests, and pyrethroid-resistant strains of the Australian carpet beetle *Anthrenocerus australis* and of the *Tinea transluens* clothes moth have been obtained in isolated domestic infestations in northern Australia, prompting the IWS to increase recommended application concentrations for that market in their Woolmark labeling program (137).

The future for application of insect-resist agents by dyebath, ca 1997, is uncertain. There are other approaches to protection of manufactured wool goods that are being investigated. (1) One is clean-up methods for dyehouse effluents (138), but liquor volumes and costs may be high and it is technically difficult to scavenge the last traces of pesticide. Disposal of the pesticide accumulated on the absorbent or flocculation materials also needs to be considered. (2) Another area under investigation involves new application methods, where little or no aqueous effluent is produced. WRONZ and some commercial equipment developers have examined, with varying levels of success, methods of treating manufactured carpets with spray, spray vacuum, foam, and dry carrier treatments, but there may be difficulties in achieving uniformity of application down to the base of the pile, and steaming may be required to achieving good penetration of the fiber to obtain optimum fastness (139,140). Other low volume applicators apply insect-resist agent at the yarn stage have shown good results (141,142), but it is necessary to obtain conditions suitable to allowing the active agent to penetrate the fiber. Where small liquor volumes are used, cleanup or incineration of the liquors may become feasible. (3) A diverse range of nonpesticidal approaches has also been studied (143). Attempts to find triggers for behaviors such as egg-laying and recognition of food were inconclusive, and of a range of metabolic inhibitors tested, only colored quinones, eg, plumbagin and juglone, showed interesting activity.

A promising, albeit difficult, route to development of specific agents for protection of wool relies on detailed investigation of the unique convergent adaptations of their gut systems that allow these otherwise unrelated moths and beetles to develop organs that enable them to digest keratin. Significant papers have followed the earlier work (144–148). One group has concentrated on the midgut proteinases, and examined the *in vitro* effects of proteinase inhibitors on the enzymes. Significant differences were found between the moth and beetle species examined (149). Another group has concentrated on the gut mechanism responsible for the maintenance of reducing conditions, and shown that cysteine lyase–desulphydrase activity is present in the gut of the clothes moth but not in the carpet beetle (150). This system converts cysteine to pyruvic acid and hydrogen sulfide, and this finding is consistent with a previous observation (151) that clothes moths, but not carpet beetles, excrete heavy metals in their diet as the insoluble sulfides. This indicates that clothes moths and carpet beetles use different mechanisms to reduce the disulfide bonds of wool.

Setting

Setting operations form an important part of the processing of wool yarn, fabric, and garments. Yarn is set in steam to stabilize twist and prevent snarling during winding and warping. Some fabrics are set after weaving to prevent the formation of distortions during wet processes such as scouring and dyeing. Normally fabrics are again flat-set near the end of the finishing routine to impart dimensional stability and to confer the required handle. Finally, garments are set in presses to form their final shape.

Three forms of set may be conferred to wool fibers. (1) Cohesive set is imparted when the fibers are dried under strain or set in steam and is lost when the fibers are relaxed in water at room temperature. (2) Temporary set is imparted at higher temperatures and is lost when the fibers are wet out in hot (70°C) water. (3) Permanent set is imparted in boiling water and in high pressure steam, and is stable to release in hot water.

Setting is a process of stress relaxation which results from rearrangement of the protein macromolecules that form the fiber. At normal regain (moisture content) and temperature, wool fibers are glass-like and when they are deformed, stress relaxation is slow. If wool is heated or wet out it becomes rubber-like and stress relaxation occurs much more rapidly. Cohesive setting occurs if the fibers are deformed under conditions where stress relaxation is high, then cooled or dried to a glassy state before they are released. The fiber will maintain its new shape until it is wet out and again becomes rubber-like.

In addition to the restrictions on their mobility caused by steric and polar interactions between chemical groups, the protein molecules in wool fibers are covalently cross-linked by disulfide bonds. Permanent setting only occurs if these disulfide bonds are also rearranged to be in equilibrium with the new shape of the fiber. Disulfide bond rearrangement occurs only at high temperature (>70°C) in wet wool and at even higher temperatures (above 100°C) in

wool at normal regain. The rearrangement is catalyzed by thiol anions in the fiber (152–154). Reducing agents which break disulfide bonds to form thiol anions catalyze the rearrangements required for permanent setting.

Depending on the conditions, wet-setting operations (called crabbing, potting, or beaming but including dyeing) impart temporary and permanent set to wool fibers. The most common form of wet-setting is continuous crabbing, an operation in which the wet fabric is sandwiched between a hot (to 160°C) roller and an impermeable belt and heated to temperatures above 100°C for up to 1 min. The amount of permanent set imparted in a continuous crab is less than that conferred in the traditional batch crabbing operation, but the continuous operation has greater productivity. The effectiveness of the operation depends on the pH of the wool (155) as well as the time and temperature of crabbing (150). The rate of permanent setting can be increased by impregnating the fibers with a reducing agent.

Steam-setting operations can impart all forms of set to wool yarn and fabric. Yarns are normally steam-set in an autoclave at temperatures up to 90°C. A vacuum pump is used to remove air from the packages and allow even penetration of the steam into the yarn packages and even setting. Fabrics are normally steam-set by winding them, interleaved with a cotton wrapper cloth, onto a perforated drum through which steam is forced. This process is called decatizing. The fabric roll can be placed in a pressure vessel so that the setting temperature can be raised above 100°C.

The amount of set imparted in steam-setting depends on the pH and moisture content of the wool as well as the time and temperature of setting. In continuous machines the conditions are mild and little permanent set is imparted. In batch machines, the wool is often set at temperature up to 136°C for 3–5 min. This imparts large amounts of permanent set to the fibers. Impregnation of the wool with reducing agents increases the rate of permanent setting. The amount of cohesive set imparted in both batch and continuous machines depends on the temperature of the fibers when the fabric is released from the wrapper cloth. The cooler the fabric, the more cohesive the set imparted.

Although designed to color the fiber, dyeing operations also impart large amounts of permanent set to wool.

In addition to normal pressing operations that impart only cohesive set to the fibers, a number of commercial processes are used to set permanent creases into wool garments (156–159). These include wet and dry pressing, both using chemical assistants, as well as autoclave setting. Autoclave setting at around 110°C is widely used to fix permanent pleats into skirts. In the SIRO-SET and Immacula processes for permanent creasing of trousers, a chemical assistant, eg, monoethanolamine sulfite, is sprayed onto the trousers before pressing.

Shrink-Resist Treatment

When a wool fabric or yarn is immersed for the first time in water, it contracts owing to a combination of reversible hygral expansion and irreversible relaxation shrinkage. Relaxation shrinkage results from the release of strains imparted during previous processing operations such as spinning, knitting, and fabric finishing (15). Felting shrinkage, characterized by the irreversible migration of individual fibers, occurs subsequently when the relaxed fabric is subject to more severe mechanical action as in laundering, but it can also occur in other operations which involve mechanical action, such as tumble drying or dry cleaning (15). Shrink-resist treatments are essentially directed at the prevention of felting shrinkage, whereas control of relaxation shrinkage requires attention to processing, particularly to fabric-finishing routines.

Terminology and Testing. The term *shrink-resistant* is preferred to "shrinkproofed," although in recent years emphasis has switched to performance-related terms such as hand-washable, and machine-washable. In Australia treatments with a high level of machine washability are called Superwash. A recent trend is to use the term *easy care*, which with knitwear generally means resistance to felting shrinkage, but may often include resistance to pilling. In woven garments easy care also includes retention of pleats or creases as well as smooth drying performance. All these terms are meaningless unless related to some testing sequence and criteria for passing the test.

Relaxation shrinkage tests usually involve a mild agitation as well as immersion in water, and then felting shrinkage is determined after more severe agitation. Hence, in practice it may be difficult to say when relaxation ceases and felting starts. The recent trend in tests for felting shrinkage is to use multiple wash test cycles in domestic-type washing machines (front loading in Europe, or top loading in the United States) and away from accelerated testing machines such as the Cubex. Many modern test methods include tumble drying. The International Wool Secretariat (IWS), as part of their Woolmark certification program, have developed test methods for machine washability together with performance criteria for different garment types and end uses. IWS tests are widely used around the world, although in the United States AATCC methods predominate.

Mechanism of Felting Shrinkage. Felting results from the presence of scales on the wool fiber surface which point away from the fiber root. The difference in friction coefficient with and against the direction of the scales results in fiber movement towards the roots under the influence of an external force. Felting of loose fibers results in entanglement, while in fabrics fiber migration inside and between yarns reduces the fabric area, ie, shrinkage occurs. The mechanisms of felting shrinkage and its prevention have been comprehensively discussed (15).

Many factors influence felting shrinkage, and they are related to three general areas: (1) fabric construction, eg, woolen vs worsted, yarn twist, etc; (2) the method of wash testing, eg, pH, temperature of wash liquor, presence of detergents, salts or lubricants, and the severity of the mechanical action of the washing machine; or (3) the properties of the fiber, eg, elasticity, fiber diameter, and fiber length. Recently felting shrinkage has been related to the glass transition temperature (160). Important points to note are that fine wools felt more readily than coarse, and that if a small amount of untreated wool is added to shrink-resist treated wool, the blend will felt as readily as untreated.

Development of Shrink-Resist Science and Technology. Felting of wool and its use in fabric finishing has been known for a very long time. Monge observed the differential friction effect in wool fibers in 1790, and Mercer chlorinated wool to improve dye affinity in printing in 1839, but chlorination was soon found to impart shrink-resistance (15,16). Chlorination was widely used during World War II to produce shrink-resistant wool socks and blankets for the military. From then on, two distinct trends in shrink-resist research and technology emerged: firstly, better control of the reaction of chlorine or other oxidants with wool, and secondly the use of synthetic polymers to bond adjacent wool fibers together. There is a vast and confusing journal and patent literature on shrink-resist treatments, and is usually difficult if not possible to assess or compare treatments owing to the use of different test fabrics and wash test methods, etc. The reader is directed instead to Reference 161, which gives a good summary up to the early 1950s, and to Reference 15, which continues to the late 1970s and has a particularly useful discussion on the mechanism of felting. Since that time the five-year International Wool Textile Research Conference papers have given useful accounts of current developments in shrink-resist technology. R&D on shrink-resist treatments continues to be directed largely to improving the performance particularly of associated properties such as handle. In addition, emphasis in recent years (ca 1997) has changed to the broader coverage of environmental aspects, particularly in the reduction or elimination of organochlorine compounds in effluents from shrink-resist processing. In future years, replacements for chlorination processes are likely to emerge.

The IWS has for some years collected statistics on wool shrink-resist treatments. As of ca 1997, about 35–40 × 10⁶ kg is treated worldwide per year. This figure may include some low level chlorination treatments of fine wools designed to prevent felting during wet processing or to improve garment appearance or dyeability, but which are generally not claimed to be machine-washable. The fact that only 5–10% of all wool is shrink-resist-treated indicates that many end uses do not require treatment. More than 90–95% of the wool that is shrink-resist-treated ends up in knitwear rather than woven products. However, the amount treated varies markedly from country to country and the type of product, with those worn next to the skin, eg, underwear, being most likely to be treated. Although wool can be shrink-resisted at various stages between loose wool and garments, IWS figures show that about 5% is treated at the loose wool stage, mainly for use in bedding products such as futons; 80% at the top stage; none as yarn; 10% as knitted garments; and about 5% as woven fabrics. Chlorine-Hercosett is by far the most important shrink-resist process in use ca 1997, accounting for at least 80% of the wool treated at the loose wool or top stage.

Industrial Shrink-Resist Treatments. A very large number of industrial treatments have been developed to allow for the requirements of different end products, processing equipment, and processing stages between loose wool and garments, but most are no longer used. Mechanistically, shrink-resist treatments are often divided into degradative and additive types. Degradative treatments aim to remove the differential friction effect of the surface scales by the use of an oxidizing agent to modify or remove the fiber scales, and in some cases a polymer is also applied to the surface to mask the scales. The additive approach prevents the migration of fibers by bonding fibers together with elastomeric polymers. Some polymer deposition treatments involve a degradative chemical pretreatment to improve the adhesion of the polymer to the fiber surface. Additive treatments must be applied after spinning (usually at the fabric stage), whereas degradative treatments can be used at any stage from loose wool through to garments.

The principal oxidizing agent used in degradative shrink-resist treatments is chlorine. Free chlorine reacts very rapidly with wool; hence it is difficult

to treat a mass of wool fibers evenly. Consequently, two different chlorination technologies have developed, firstly continuous top or loose wool treatments in which the wool is reacted with free chlorine for a short time (<30 s), and secondly, slow (5–30-min) batch treatments with a less reactive chlorinating agent such as DCCA (N,N' -dichloroisocyanuric acid). These batch treatments are mainly used on garments and fabrics but can be applied to loose wool, or tops. Generally in both continuous and batch treatments a reactive polymer, usually cationic, is applied. The only other oxidizing agent used to any extent is permonosulfuric acid, HOOSO_3H , usually in the form of a potassium triple salt containing potassium sulfate and bisulfate.

The features and chemistry of chlorination shrink-resist treatments are best illustrated by consideration of the chlorine-Hercosett process, which accounts for more than 80% of the shrink-resist-treated wool (Fig. 7). This process uses special dedicated plants, of which there are about 40 around the world. A web of 30–40 parallel slivers (usually 20–30 kg km) is passed at 5–10 m min through a series of bowls, where water and chemicals are forced through the web of wool. The bowls are separated by squeeze rollers, and finally the wool is dried. Overall production rates range from 200 to 500 kg h. The first stage involves chlorination, usually with elemental chlorine dissolved in water; alternatively, sodium hypochlorite can be acidified with sulfuric acid. Chlorination is the most critical step of the process, and to achieve uniform chlorination of the web, modern plants chill the chlorine solution to 10–12°C and use special devices, eg, those produced by the Kroy or Fleissner companies, to rapidly contact the web with a large volume of chlorine solution. The second stage is termed neutralization and involves treatment with sodium sulfite; the third is a treatment with a cationic polymer; and the fourth is a treatment with a cationic softener and a lubricant to facilitate further processing. Some of these stages may be separated by water rinses, or the neutralization may be divided into a sodium bisulfite stage followed by an alkaline (pH 9) protein strip stage. The reaction of chlorine with wool is complex, but the principal effect of the combination of the chlorination and neutralization stages is to remove surface lipid from the fiber and hence increase the surface energy, shift the isoelectric point of the fiber to a higher pH by conversion of cystine disulfides into cysteic acid sulfonic acid groups, and to solubilize and remove some of the surface protein. These changes allow cationic polymer to deposit on the fiber surface in the third stage (Fig. 8), usually Hercosett (Hercules Inc.), a polyamide epichlorohydrin resin which cross-links during drying. Virtually all of the shrink-resistance is imparted by the chlorination, and the function of the Hercosett is to make up for any weight loss during chlorination, prevent the further loss of anionic soluble protein during subsequent dyeing operations, and to improve the shrink-resistance of any insufficiently chlorinated fibers. In batch treatments, one treatment bath is used and after an initial pH neutralization and scour, the chlorinating agent, neutralization chemicals, and polymers are added successively after suitable delays for reaction.



Fig. 7. Electron micrographs showing (a) an untreated wool fiber and (b) a wool fiber treated with chlorine-Hercosett.



Fig. 8. Electron micrograph of Merino wool fibers in a fabric that have been treated with a typical shrink-resistance polymer, showing fiber–fiber bond formation.

The principal additive shrink-resist treatment uses the polymer Synthappret BAP (Bayer AG) which is a polypropylene oxide polyurethane containing reactive carbamoyl sulfonates (or isocyanate bisulfite adduct groups, $\text{-NHCO}_2\text{SO}_3\text{-Na}^+$). An aqueous solution of this polymer is padded onto woven fabrics, which are immediately dried. Other polymers may be applied at the same time to modify the handle.

Durable Press Wool. Durable (or permanent) press describes woven garments that can be machine washed and tumble dried, and retain any imparted creases or pleats while flat areas dry smoothly with minimum wrinkling, so that garments can be worn again with minimum or no ironing. Easy care is replacing the terms durable and permanent press. An easy care process requires that the garments, usually trousers and, to a lesser extent, skirts, are made from shrink-resistant fabrics and are also subject to a setting process which imparts set stable to machine washing.

To produce easy care wool–polyester blend garments, heat setting of the polyester component will impart set stability to repeated machine washing. A minimum of 20–30% polyester is needed for adequate stability. A fabric shrink-resist treatment such with Synthappret BAP may be necessary to give the required shrink-resistance for easy care performance. If the polyester content is increased, particularly above 50%, and a suitable fabric construction is used, heat setting alone will give easy care performance.

It is more difficult to produce pure wool easy care garments. The problem is to obtain adequate durability of the setting rather than obtaining shrink-resistance, because some wool setting processes are not stable to machine washing. As of ca 1997, the best approach is to combine the Synthappret BAP fabric shrink-resist process with Siroset garment setting.

Flame-Resistance Treatment. Wool is naturally a very flame-resistant fiber, and in most applications, no special treatment is needed. However, consumer demand for higher standards of flammability in, for example, aviation furnishings, drapes, and carpets, has led to the development of treatments for improving this property. The treatments evolved from the observation that chrome mordanting of wool markedly improves its flame resistance (162). However, this approach did not prove practical for all situations because of the discoloration of undyed wool by the chrome mordant. Subsequently, it was shown that titanium or zirconium compounds give similar results (163), but the zirconium compounds did not share the discoloration disadvantage. The process, ie, the IWS Zirpro process, has been used extensively for wool in aircraft interiors to meet FAA requirements and for curtains and carpets in public buildings and similar areas where high flame resistance is required by legislation (164,165) (see FLAME RETARDANTS FOR TEXTILES). Shrink-resistance processes compatible with Zirpro-treated wool have also been developed (166).

Wool Grease

In wool scouring, the contaminants on the wool, mainly grease, dirt, suint, and protein material, are washed off the fiber and remain in the wastewaters either in emulsions or suspension (grease, dirt, protein) or in solution (suint). Centrifugal extraction of the wastewaters produces a grease contaminated with detergent and suint. This product is called wool grease.

Lanolin is wool grease that has been refined to lighten its color and reduce its odor and free fatty acid content. Wool wax is the pure lipid material of the fleece, yellow to pale brown in color and extractable with the usual fat solvents such as diethyl ether and chloroform. Wool grease is a mixture of compounds that are classed as waxes. However, it does not have the physical characteristics usually displayed by waxes. It is soft and slightly sticky, with a greasy appearance. Some centrifugally recovered greases are light buff or ivory in color and practically odorless. Those recovered by other methods contain

dark-colored and odorous impurities as well as free fatty acids. Grease recovered by acid-cracking of scour wastewaters can be used in lanolin production, but the centrifugally recovered products are more suited to the alkali refining, bleaching, and deodorizing required in preparing pharmaceutical-grade lanolin. Table 8 gives some physical and chemical data for wool wax.

Table 8. Some Physical and Chemical Data for Wool Wax^a

Property	Value
density, at 15°C	0.94–0.97
refractive index, at 40°C	1.48
mp, °C	35–40
free acid content, %	4–10
free alcohol content, %	1–3
iodine value (Wijs)	13–30
saponification value	95–120
mol wt ^b	790–880
fatty acids, %	50–55
alcohols, %	50–45
acids	
mp, °C	40–45
iodine value (Wijs)	10–20
mean mol wt	330
alcohols	
mp, °C	55–65
iodine value, Wijs	40–50
mean mol wt	370

^a Ref. 167.
^b Rast method, in phenyl salicylate.

Chemical Composition. Wool wax is a complex mixture of esters of water-soluble alcohols (168) and higher fatty acids (169) with a small proportion (ca 0.5%) of hydrocarbons (170). A substantial effort has been made to identify the various components, but results are complicated by the fact that different workers use wool waxes from different sources and employ different analytical techniques. Nevertheless, significant progress has been made, and it is possible to give approximate percentages of the various components. The wool-wax acids (Table 9) are predominantly alkanoic, α -hydroxy, and ω -hydroxy acids. Each group contains normal, iso, and anteiso series of various chain length, and nearly all the acids are saturated.

Table 9. Summary of the Average Composition of Wool–Wax Acids^a

Acids ^b	Chain length ^c	Wool–wax acids, %
normal acids	C ₈ –C ₃₈	10
iso-acids	C ₈ –C ₄₀	22
anteiso-acids	C ₇ –C ₄₁	28
normal α -hydroxy acids	C ₁₀ –C ₃₂	17
iso- α -hydroxy acids	C ₁₂ –C ₃₄	9
anteiso- α -hydroxy acids	C ₁₁ –C ₃₃	3
normal ω -hydroxy acids	C ₂₂ –C ₃	3
iso- ω -hydroxy acids	C ₂₂ –C ₃	0.5
anteiso- ω -hydroxy acids	C ₂₂ –C ₃	1
polyhydroxy acids		4.5 ^d
unsaturated acids		2 ^d

^a Ref. 169.
^b In the iso series, a branching methyl group is attached at the penultimate position, and in the anteiso series, at the antepenultimate position.
^c Ref. 171.
^d Tentative data.

The alcohol fraction is likewise a complex mixture of both aliphatic and cyclic compounds (Table 10). The principal components are cholesterol (34%), and lanosterol and dihydrolanosterol (38%). The aliphatic alcohols account for about 22% of the unsaponifiable products. Sixty-nine components of aliphatic alcohols had been reported up to 1974 (latest reported work as of ca 1997). The hydrocarbons (ca 0.5%) show structural similarity to the wool–wax acids or aliphatic alcohols and contain highly branched alkanes as well as cycloalkanes.

Table 10. Summary of the Average Composition of Wool–Wax Alcohols^a

Alcohol ^b	Chain length ^c	Approximate % of wool-wax alcohols
normal monoalcohols	C ₁₄ –C ₃₄	2
iso-monoalcohols	C ₁₄ –C ₃	
anteiso-monoalcohols	C ₁₇ –C ₃₅	13
normal alkan-1,2-diols	C ₁₂ –C ₂₅	1
iso-alkan-1,2-diols	C ₁₄ –C ₃₀	

anteiso alkan-1,2-diols	C ₁₅ –C ₂₀	6
<i>Total</i>		22
cholesterol		34
lanosterol		
dihydrolanosterol		38
<i>Total</i>		72
hydrocarbons		1
autooxidation products		5
undetermined		

^a Refs. 168, 171–174.
^b See footnote ^b, Table 9.
^c Ref. 171.

Wool-Grease Recovery. Current systems and research on wool-grease recovery have been reviewed comprehensively (175). The principal recovery process in use involves centrifuging in a cream-separator type of centrifuge modified by the addition of peripheral jets or other mechanical devices to remove dirt. Liquor is usually withdrawn from the second bowl of a five-bowl scouring plant, heated to about 90°C, and passed into a large settling tank to remove the heaviest dirt particles. The hot emulsion is then passed through a centrifuge designed to remove suspended solids, reheated, and run through the cream-separator type of centrifuge. This separates most of the grease remaining in the liquors. The liquor is cooled to about 50°C and returned to the scouring plant or held in a reserve tank for later use.

With the introduction of scouring with nonionic detergents instead of soap and soda, grease-recovery rates increased (176); however, the quality of the grease decreased slightly because of increased recovery of the grease from the tip of the fiber (oxidized grease) (177). The recovery of grease by centrifuging has been optimized for normal scouring conditions in the WRONZ Comprehensive Scouring System (66), and the Lo-flo system of scouring (63) yields grease recoveries of about 85% of the grease originally on the wool, although the quality of the grease is again reduced.

Other techniques aimed at improving grease recovery (and often attempting also to improve the scouring process itself) have included solvent degreasing of the wool (52,53), solvent extraction of the liquor or sludge (178), aeration (179,180), and physical and chemical destabilization (175).

Grease Refining and Fractionation. Lanolin to be used in pharmaceuticals and cosmetics must conform to strict requirements of purity, such as those in the U.S. and British Pharmacopoeias (181,182). These include specifications for the maximum allowable content of free fatty acids, moisture, ash, and free chloride. Lanolin intended for certain dermatological applications may have to meet further specifications in relation to free-alcohol and detergent contents (183,184).

The refining process most commonly used involves treatment with hot aqueous alkali to convert free fatty acids to soaps, followed by bleaching, usually with hydrogen peroxide, although sodium chlorite, sodium hypochlorite, and ozone have also been used. Other techniques include distillation, steam stripping, neutralization by alkali, liquid thermal diffusion, and the use of active adsorbents, eg, charcoal and bentonite, and solvent fractionation (175).

Uses of Wool Grease. The uses of wool grease, lanolin, and lanolin derivatives are many, ranging from pharmaceuticals and cosmetics (qv) to printing inks (qv), rust preventatives, and lubricants (see LUBRICATION AND LUBRICANTS).

In pharmaceutical uses, the general inertness of lanolin and its derivatives, together with their ease of emulsification (185), have been important criteria. These properties are also important in cosmetics, but with emphasis also on the ability to absorb large quantities of water or, after suitable modification, to stabilize oil-in-water emulsions.

The anticorrosion properties of lanolin have been utilized over a considerable period, and the product has the status of a temporary corrosion inhibitor (186) (see CORROSION AND CORROSION CONTROL). There is probably potential for greater use of the poorer grades of lanolin for long-term storage and protection of machine parts. Other uses have been reviewed in detail (182).

ACKNOWLEDGMENT

Much of the work supporting the findings in this article was carried out at the CSIRO Division of Wool Technology with its principal laboratories in Geelong, Australia, and at the IWS Development Centre in Ilkley, U.K. Funding for this research came from woolgrowers through the IWS, and from the Australian Government through CSIRO.

BIBLIOGRAPHY

"Wool" in *ECT* 1st ed., Vol. 15, pp. 103–134, by J. Menkart, Textile Research Institute; in *ECT* 2nd ed., Vol. 22, pp. 387–418, by M. Lipson, Division of Textile Industry, CSIRO; in *ECT* 3rd ed., Vol. 24, pp. 612–644, by W. S. Boston, CSIRO.

1. *Textile Terms and Definitions*, 10th ed., Textile Institute, Manchester, England, 1996.
2. International Wool Textile Organization (IWTO) Wool Statistics 1994–1995.
3. M. Lipson and U. A. F. Black, *J. Roy. Soc. New South Wales* **78**, 84 (1945).
4. R. B. Sweeten, *J. Text. Inst. Trans.* **40**, T727 (1949).
5. *CSIRO Woolspec 94, Seminar Proceedings*, CSIRO, Sydney, Australia, 1994.
6. G. A. Camaby and A. J. McKinnon, *Australas. Text.* **2**(2), 11 (1982).
7. *Objective Measurement of Wool in Australia*, Australian Wool Corp., Melbourne, Australia, 1973.
8. B. H. Mackay, *Proc. Int. Wool Text. Res. Conf. (Pretoria)* **1**, 59 (1980).
9. Core Test Regulations, Test Methods, IWTO 19-76, 28-75, 3-73, 20-69, International Wool Textile Organisation, London, 1968.
10. Australian Standard 1363-1976, Standards Association of Australia, Melbourne, Australia, 1976.
11. M. H. Hardy and A. G. Lyne, *Austral. J. Biol. Sci.* **9**, 423 (1965).
12. R. E. Chapman and K. A. Ward, in J. L. Black and P. J. Reis, eds., *Physiological and Environmental Limitations to Wool Growth*, University of New England Publishing Unit, Armidale, Pa., 1979, pp. 193–208.
13. R. D. B. Fraser, T. P. MacRae, and G. E. Rogers, *Keratins, Their Composition, Structure and Biosynthesis*, Charles C. Thomas, Springfield, Ill., 1972.
14. L. N. Jones, T. J. Horr, and I. J. Kaplin, *Micron* **25**, 589–595 (1994).
15. K. R. Makinson, *Wool Shrinkproofing*, Marcel Dekker, Inc., New York, 1979, p. 147.
16. R. C. Marshall, D. F. G. Orwin, and J. M. Gillespie, *Electron Microsc. Revs.* **4**, 47–83 (1991).
17. I. J. Kaplin and K. J. Whiteley, *Proc. 7th Int. Wool Text. Res. Conf. (Tokyo)* **1**, 95 (1985).
18. J. D. Leeder, *Wool Sci. Rev.* **63**, 3–35 (1986).
19. I. C. Watt and R. L. D'Arcy, *J. Text. Inst.* **70**, 298 (1979).
20. G. R. S. Naylor, *Wool Tech. Sheep Breed.* **40**, 14 (1992).
21. M. Feughelman, *Text. Res. J.* **59**, 739 (1989).

22. A. R. Haly and J. W. Snaith, *Text. Res. J.* **37**, 898 (1967).
23. F. J. Wortmann, B. J. Rigby, and D. G. Phillips, *Text. Res. J.* **54**, 6 (1984).
24. D. G. Phillips, *Text. Res. J.* **55**, 171 (1985).
25. S. Rosenbaum, *J. Polym. Sci.* **31C**, 45–55 (1970).
26. F.-J. Wortmann, *Melliand Textilber.* **66**, 78–80 (1985).
27. M. Feughelman, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, Vol. 8, 1987, p. 566.
28. J. W. S. Hearle and M. Susutoglu, *Proc. 7th Int. Wool Text. Res. Conf. (Tokyo)* **1**, 214 (1985).
29. F. J. Wortmann and H. Zahn, *Text. Res. J.* **64**, 737 (1994).
30. T. W. Mitchell and M. Feughelman, *Text. Res. J.* **30**, 662 (1960).
31. J. B. Speakman, *J. Text. Inst.* **18**, T431 (1927).
32. M. Feughelman and M. S. Robinson, *Text. Res. J.* **41**, 469 (1971).
33. *Wool Research*, Vol. 2, Wool Industries Research Association, Leeds, U.K., 1955.
34. W. E. Morton and J. W. S. Hearle, *Physical Properties of Textile Fibres*, 3rd ed., The Textile Institute, Manchester, U.K., 1993.
35. B. Milligan, L. A. Holt, and J. B. Caldwell, *Appl. Polymer Symp.* **18**, 113 (1971).
36. M. Cole, J. C. Flecher, K. L. Gardner, and M. C. Cordield, *Appl. Polymer Symp.* **18**, 147 (1971).
37. L. A. Holt, B. Milligan, and L. J. Wolfram, *Text. Res. J.* **44**, 846 (1974).
38. J. M. Gillespie, in R. D. Goldman and P. M. Steinert, eds., *Cellular and Molecular Biology of Intermediate Filaments*, Plenum Press, New York, 1990, p. 95.
39. R. D. B. Fraser, T. P. MacRae, and D. A. D. Parry, in Ref. 38, p. 205.
40. J. A. Rippon, in D. M. Lewis, ed., *Wool Dyeing*, Society of Dyers and Colourists, Bradford, U.K., 1992, p. 1.
41. A. D. Parry, in Ref. 38, p. 175.
42. L. M. Dowling, K. F. Ley, and A. M. Pearce, *Proc. Int. Wool Text. Res. Conf. (Christchurch)* **1**, 205 (1990).
43. J. F. Conway, R. D. B. Fraser, T. P. Macrae, and D. A. D. Parry, in G. E. Rogers, P. J. Reis, K. A. Ward, and R. C. Marshall, eds., *The Biology of Wool and Hair*, Chapman and Hall, London, 1988, p. 127.
44. D. E. Rivett, *Wool Sci. Rev.* **67**, 1 (1991).
45. J. Herdling and H. Zahn, *Proc. 7th Int. Wool Text. Res. Conf. (Tokyo)* **1**, 181 (1985).
46. A. Schwan, A. J. Herrling, and H. Zahn, *Colloid Polymer Sci.* **264**, 171 (1986).
47. A. Breasley and J. A. Iredale, *The Worsted Industry: An Account of the Worsted Industry and Its Processes from Fibre to Fabric*, Wira, Leeds, U.K., 1980.
48. R. C. Stewart, *Wool Sci. Rev.*, pp. 16–80 (Dec. 1987); R. C. Stewart and R. C. Jamieson, *ibid.*, pp. 16–24; J. R. Christie, *ibid.*, pp. 25–43; B. V. Harrowfield, *ibid.*, pp. 44–80.
49. G. F. Wood, *Text. Prog.* 12(1), (1982).
50. *Text. J. Aust.* **44**(6), 34 (1969).
51. *Text. Rec.*, 69 (Jan. 1964).
52. CSIRO Division of Textile Industry Report No. G. 10, CSIRO Division of Textile Industry, Geelong, Australia, 1960.
53. J. Brach, *Wool Sci. Rev.* (36), 38 (May 1969).
54. B. Robinson, C. P. S. Lee, and T. Shaw, *Technical Information Letter, TIL/ET-14*, International Wool Secretariat, Ilkley, U.K., Jan. 1993.
55. A. S. C. Lawrence, *Nature* **183**, 1491 (1959).
56. R. P. Harker, *J. Text. Inst. Trans.* **50**, T189 (1959).
57. C. A. Anderson, *J. Text. Inst.* **73**, 289 (1982).
58. C. A. Anderson, *Australas. Text.* **2**(5), 28 (1982).
59. B. O. Bateup, *Wool Scouring and Worsted Carding: New Approaches*, CSIRO Division of Wool Technology Symposium, Geelong, Australia, Nov. 1986.
60. C. A. Anderson and J. R. Christoe, *Textile Res. J.* **52**(12), 741 (1983).
61. B. Robinson, C. P. S. Lee, and T. Shaw, *Technical Information Letter, TIL/ET-14*, International Wool Secretariat, Jan. 1993; J. R. Christoe and B. O. Bateup, CSIRO Division of Wool Technology Report, G70, CSIRO Division of Wool Technology, Geelong, Australia, 1992.
62. C. A. Anderson, J. R. Christoe, A. J. C. Pearson, J. J. Warner, and G. F. Wood, *Developments in Scouring and Carbonising 1980*, Disc 80, Australian Wool Corp., Melbourne, Australia, 1980, p. 1.
63. G. F. Wood, A. J. C. Pearson, and J. R. Christoe, CSIRO Division of Textile Industry Report No. G 39, CSIRO Division of Textile Industry, Geelong, Australia, 1980.
64. K. Peverill, K. Reed, I. M. Russell, J. Tisdall, *Opportunities for Wool Processing, A Report for the Government of Victoria*, Business Victoria, Melbourne, Australia, 1995.
65. J. R. Christoe and B. O. Bateup, *Proceedings of Third Asian Textile Conference*, Hong Kong Polytechnic University, Hong Kong, 1995.
66. R. G. Stewart, G. V. Barker, P. E. Chisnall, and J. L. Hoare, WRONZ Report No. 25, Wool Research Organisation of New Zealand, Inc., Christchurch, New Zealand, 1974.
67. R. G. Stewart, WRONZ Report No. 27, Wool Research Organisation of New Zealand, Inc., Christchurch, New Zealand, 1974.
68. R. G. Stewart, P. E. Chisnall, J. M. Flynn, and R. G. Jamieson, WRONZ Communication No. 36, Wool Research Organisation of New Zealand, Inc., Christchurch, New Zealand, 1975.
69. N. C. Gee, Shoddy and Mungo Manufacture, Emmott & Co., Bradford, U.K., 1950.
70. H. K. Rouette and G. K. Kittan, *Text. Praxis Int.* **36**, 784 (1981).
71. W. G. Crewther, *Proc. Int. Wool Textile Res. Conf. (Australia)* **E**, 408 (1955).
72. W. G. Crewther and T. A. Pressley, *Text. Res. J.* **28**, 67 (1958).
73. A. E. Davis, A. J. Johnson, and L. R. Mizell, *Text. Res. J.* **31**, 825 (1961).
74. M. S. Nossler, M. Chaikin, and A. J. Datyner, *J. Text. Inst.* **62**, 677 (1971).
75. M. S. Nossler and M. Chaikin, *Proc. Int. Wool Text. Res. Conf. (Aachen)* **IV**, 136 (1975).
76. H. J. Katz, D. E. A. Plate, and G. W. Wood, *Wool Sci. Rev.* (45), 28 (1973).
77. M. A. White, *Text. Prog.* **13**(2), (1983), p. 1.
78. H. K. Rouette and A. Ohm, *Melliand Textilber.* **62**, 583 (1981).
79. H. K. Rouette, A. Gotz, and A. Ohm, *Int. Text. Bull. Dye. Print. Finish.* (3), 211 (1981).
80. M. A. White, in *Proceedings of a Conference on Recent Developments in Wool and Wool Blend Processing*, CSIRO Division of Textile Industry, Geelong, Australia, 1983, p. 20.
81. L. A. Allen and co-workers, *Proc. 8th Quinq. Wool Text. Res. Conf. (Christchurch)* **III**, 412 (1990).
82. P. R. Lamb and co-workers, DWI Report 111, Aachener Textil-Tagung, 1993, p. 63.
83. Report to Topmaker No. 11, Lubrication for Improved Worsted Top, Report No. G65, Oct. 1989.
84. G. Fisher, *Wool Rec.* **154**(3614), 23 (Dec. 1995).
85. *Proc. 1st China Int. Wool Text. Conf.* **1**, 122 (1994).
86. K. R. Atkinson, *J. Text. Inst.* **83**(4), 591–602 (1992).
87. J. A. Rippon in Ref. 40, Chapt. 1.
88. C. L. Bird, *The Theory and Practice of Wool Dyeing*, Society of Dyers and Colourists, Bradford, U.K., 1972.
89. L. Peters in C. L. Bird and W. S. Boston, eds., *The Theory of Coloration of Textiles*, Dyers' Company Publications Trust, Bradford, U.K., 1975, p. 163.

90. R. H. Peters, *Textiles Chemistry*, Vol. III, Elsevier Publishing, Amsterdam, the Netherlands, 1975, p. 203.
91. J. Meybeck and P. Galafassi, *Proc. 4th Int. Wool Text. Res. Conf. (San Francisco)*, **Pt. 1**, 463 (1970).
92. M. T. Pailthorpe in Ref. 40, Chapt. 2.
93. D. M. Lewis, *Rev. Prog. Color* **8**, 10 (1977).
94. J. A. Bone, J. Shore, and J. Park, *J. Soc. Dyers Colour*, **104**, 12 (1988).
95. J. D. Leeder, J. A. Rippon, F. E. Rothery, and I. W. Stapleton, *Proc. 7th Int. Wool Text. Res. Conf. (Tokyo)*, **5**, 14 (1985).
96. J. D. Leeder, L. A. Holt, J. A. Rippon, and I. W. Stapleton, *Proc. 8th Int. Wool Text. Res. Conf. (Christchurch)*, **4**, 227 (1990).
97. P. A. Duffield, in Ref. 40, Chapt. 6.
98. S. M. Burkinshaw, in Ref. 40, Chapt. 7.
99. D. M. Lewis, in Ref. 40, Chapt. 8.
100. F. W. Marriott, in Ref. 40, Chapt. 5.
101. P. R. Brady and J. A. Rippon, *J. Soc. Dyers Colour*, **108**, 114 (1992).
102. J. A. Rippon, F. J. Harrigan, and A. R. Tilson, *Proc. 9th Int. Wool Text. Res. Conf. (Biella)*, **III**, 122 (1995).
103. S. M. Smith and co-workers, *ibid.*, 113.
104. V. Bell, *J. Soc. Dyers Colour*, **104**, 159–172 (1988); V. Bell, in Ref. 40, p. 332.
105. J. Ryu, T. Wakida, and T. Takagishi, *Text. Res. J.* **61**, 595–601 (1991); K. M. Byrne, A. Barwick, D. Gerstenberg, and V. Bell, *DWT Rep.* **109**, 589–599 (1992).
106. W. Rakowski, *Plasma Treatment of Wool*, PhD. dissertation, Polish Textile Institute, Lodz, 1991; *idem*, *Pr. Inst. Wlok.* **36–37**, 184–201 (1987); K. M. Byrne, *Finishing of Wool using Plasma Technology*, IWS Technical Information Bulletin CPB 109, International Wool Secretariat, Ilkley, U.K., (1991).
107. K. R. Millington, *Proc. Int. Wool Text. Conf. (Biella)* **I**, 86 (1995); K. R. Millington, *Proc. Radtech '96 North America Conf. (Nashville)*, 481 (1996).
108. D. Lewis, in Ref. 40, p. 112; P. A. Duffield, *IWS Review of Bleaching*, 1987.
109. P. A. Duffield, *IWS Technical Information Bulletin DPB-31*, 1989.
110. K. M. Thompson, W. P. Griffith, and M. Spiro, *J. Chem. Soc., Chem. Comm.* **21**, 1600 (1992).
111. A. W. Karunditu and co-workers, *Text. Res. J.* **64**, 570 (1994).
112. J. Cegarr, J. Gacen, D. Cayuela, and M. C. Riva, *J. Soc. Dyers Colour*, **110**, 308 (1994).
113. *IWS Process Devel. Bull.* **13** (Apr.–May 1994).
114. A. Bereck, *Rev. Prog. Colorat.* **24**, 17 (1994).
115. W. N. Marmer, J. M. Cardamone, and M. Arifoglu, *Text. Chem. Colorist* **26**, 19 (1994).
116. C. H. Nicholls, *Appl. Sci.* **1**, 125–144 (1980); B. Milligan, *Proc. Int. Wool Text. Conf. Pretoria*, **V**, 167–181 (1980).
117. I. H. Leaver and J. F. K. Wilshire, *Chem. Austral.*, 174 (1990).
118. K. R. Millington, *Proc. Int. Wool Text. Conf. (Biella)* **III**, 174 (1990).
119. R. Levene, *Text. Res. J.* **57**, 298 (1987).
120. G. Robinson and E. Nielsen, *Tineid Genera of Australia: Lepidoptera*. CSIRO Publications, Canberra, Australia, 1993.
121. D. E. Rivett and co-workers, *Proceedings of the 8th International Wool Textile Research Conference*, Vol. IV, WRONZ, Chistchurch, New Zealand, p. 548.
122. R. J. Mayfield, *Text. Prog.* **11**(4), (1982); D. M. Lewis and T. Shaw, *Rev. Prog. Colorat.* **17**, 86–94 (1987).
123. I. Russell, *Combined Insect Resist and Rot-resist Treatments for Wool Insulation*, CSIRO Division of Wool Technology Report G69, May 1992.
124. T. Zabel, J. Seager, and S. D. Oakley, *Proposed Environmental Quality Standards For List 2 Substances in Water–Mothproofing Agents*, Report TR261, Water Research Centre, 1988.
125. *J. Chromatog. A*, **707**, 293–302 (1995).
126. T. Shaw, in *Proc. 8th Int. Wool Text. Res. Conf.*, **IV** 533, (Feb. 1990).
127. R. J. Mayfield, *Text. Prog.* **11**(4) 1, (1982).
128. D. M. Lewis and T. Shaw, *Rev. Prog. Colorat.* **17**, 86–94 (1987).
129. F. W. Jones, R. J. Mayfield, and G. J. O'Loughlin, *Pesticide Sci.* **13**, 602 (1982).
130. R. W. Moncrieff, *Mothproofing*, Billing and Sons Ltd., London, U.K., 1950, p. 70.
131. J. M. Wimbush, *J. Soc. Dyers Colour*, **104**, 37 (1988).
132. J. M. Wimbush, *J. Soc. Dyers Colour*, **106**, 241 (1990).
133. Ref. 121, p. 606.
134. Haas, in Ref. 121, p. 558, and RH 5992 (high lepidoptera activity).
135. R. M. Mayfield and I. M. Russell, *Text. Inst. Industr.* **15**, 419–420 (1977).
136. R. M. Mayfield, *J. Text. Inst.* **70**(9), 385–389 (1979).
137. Woolmark Standard Specification, E 10, International Wool Secretariat, Ilkley, U.K., 1995.
138. P. I. Norman and R. Seddon, *J. Soc. Dyers Colour*, **107**, 215 (1991).
139. P. E. Ingham and C. K. Rowan, in Ref. 21, p. 578.
140. *Wool Rec.*, 25 (Mar. 1995).
141. D. Allenach, NZ Conf Vol IV, 568.
142. D. Allenach, *Wool Rec.*, 35 (Mar. 1992).
143. D. E. Rivett and co-workers, in Ref. 121, p. 548.
144. D. F. Waterhouse, *Aust. J. Sci. Res.* **B5**, 168 (1952).
145. R. F. Powning and H. Irzkiewicz, *Aust. J. Biol. Sci.* **13**, 59 (1960).
146. C. W. Ward, *Austral. J. Biol. Sci.* **28**, 1 (1975).
147. T. Yoshimura and co-workers, *Insect Biochem.* **18**, 771 (1988).
148. J. E. Baker and S. M. Woo, *Comp. Biochem. Physiol.* **69B**, 189 (1981).
149. J. T. Christeller, N. P. Markwick, and E. P. J. Burgess, *Arch. Insect. Biochem. Physiol.* **25**, 159 (1994).
150. C. P. Robinson, S. Ciccotosto, and L. G. Sparrow, *J. Text. Inst.* **84**, 39 (1993).
151. D. F. Waterhouse, *Austral. J. Sci. Res.* **B5**, 168 (1952).
152. J. B. Caldwell, S. J. Leach, S. J. Meschers, and B. Milligan, *Text. Res. J.* **34**, 627 (1964).
153. H. D. Weigmann, L. Rebenfeld, and C. Dansizer, *Proc. Int. Wool Text. Res. Conf. (Paris)* **II**, 319 (1965).
154. W. G. Crewther, *J. Soc. Dyers Colour*, **82**, 54 (1966).
155. Ref. 17, p. 294.
156. M. A. White, CSIRO Division of Textile Industry Report No. G44, CSIRO Division of Textile Industry, Geelong, Australia, 1982.
157. A. J. Farnworth, *Am. Dyest. Rep.* **49**, 996 (1960).
158. *Bull. Cent. Text. Contr. Rech. Sci. (Roubaix)* **43**, 39 (1959).
159. J. R. Cook and J. Delmenico, *Proc. Int. Wool Text. Res. Con. (Paris)* **III**, 419 (1965).
160. A. P. Pierdot, "Influence of Glass Transition on the Felting Shrinkage of Wool Fabrics," *Text. Res. J.* (in press May 1997).
161. R. W. Moncrieff, *Wool Shrinkage and Its Prevention*, National Trade Press, London, 1953.
162. M. J. Koroskys, *Am. Dyest. Rep.* **60**(5), 48 (1971).
163. L. Benisek, *Text. Manuf.* **99**, 36 (Jan.–Feb. 1972); *Melliand Textilber.* **53**, 931 (1972).

164. L. Benisek, *J. Text. Inst.* **65**, 102, 140 (1974).
165. P. Gordon and L. Stephens, *J. Soc. Dyers Colour.* **90**, 239 (1974).
166. L. Benisek, G. Edmondson, and B. Greenwood, *Text. Inst. Ind.* **14**, 344 (1976).
167. E. V. Truter, *Wool Wax*, Cleaver-Hume Press Ltd., London, 1956, p. 32.
168. K. Motiuk, *J. Am. Oil Chem. Soc.* **56**, 651 (1979).
169. *Ibid.*, **56**, 91 (1979).
170. *Ibid.*, **57**(4), 145 (1980).
171. F. Fawaz, C. Miet, and F. Puisieux, *Ann. Pharm. Fr.* **31**(1), 63 (1973).
172. K. E. Murray and R. Schoenfeld, *Austral. J. Chem.* **8**, 424 (1955).
173. D. H. S. Horn, *J. Sci. Food Agric.* **9**, 632 (1958).
174. D. T. Downing, Z. H. Krantz, and K. E. Murray, *Austral. J. Chem.* **13**, 80 (1960).
175. R. G. Stewart and L. F. Story, WRONZ Technical Paper No. 4, Wool Research Organisation of New Zealand, Inc., Christchurch, New Zealand, 1980.
176. R. E. Wolfrom, *Am. Dyest. Rep.* **43**, P372 (1954).
177. C. A. Anderson and G. F. Wood, *Nature* **193**, 742 (1962).
178. C. Moxhet, *Ind. Text. Beige* **7/8**, 51 (1971); **9**, 83 (1971).
179. L. F. Evans and W. E. Ewers, *Austral. J. Appl. Sci.* **4**, 552 (1953).
180. **USSR Pat. 104,212** (Dec. 1956), A. S. Salin and co-workers.
181. *The United States Pharmacopoeia XX* (USP XX-NF XV), The United States Pharmacopeial Convention, Inc., Rockville, Md., 1980.
182. *British Pharmacopoeia*, Her Majesty's Stationary Office, London, 1980.
183. E. W. Clark, E. Cronin, and D. S. Wilkinson, *Contact Dermatit.* **3**, 69 (1977).
184. R. Elder, *J. Environ. Path. Toxicol.* **4**(4), 63 (1980).
185. A. Castillo and J. M. Sure, *Ars Pharm.* **9**, 367 (1968).
186. **U.S. Pat. 2,473,614** (June 21, 1949), E. Snyder (to American Chemical Paint Co.).

General References

Refs. 17, 51, 52, 53, and 176 are also general references.
Wool Sci. Rev., International Wool Secretariat, London; published on an irregular basis.
P. Alexander and R. F. Hudson in C. Earland, ed., *Wool: Its Chemistry and Physics*, 2nd ed., Chapman & Hall Ltd., London, 1963.
W. J. Onions, *Wool: An Introduction to its Properties, Varieties, Uses and Production*, Ernest Benn Ltd., London, 1962.
R. S. Asquith, ed., *Chemistry of Natural Protein Fibers*, Plenum Press, New York and London, 1977; R. G. Stewart and L. F. Story, *Wool Grease: A Review of Its Recovery and Utilisation*, Wool Research Organisation of N.Z. (Inc.), Christchurch, 1980 (comprehensive review of the literature from 1947 to mid-1979).
Papers, 1st International Wool Textile Research Conference, Australia, CSIRO, Melbourne, Australia, 1955.
Papers, 2nd International Wool Textile Research Conference, Textile Institute, Harrowgate, U.K.
J. Text. Inst. Trans. **51**, T489 (1960).
Papers, 3rd International Wool Textile Research Conference, Paris, l'Institut Textile de France, Paris, 1965.
L. Rebenfeld, ed., *Proceedings of the 4th International Wool Textile Research Conference, Applied Polymer Symposium No. 18*, International Science Publishers, a division of John Wiley & Sons, Inc., New York, 1971.
H. Ziegler, ed., *Proceedings of the 5th International Wool Textile Research Conference*, Deutsches Wollforschungsinstitut an der Technischen Hochschule, Aachen, Germany, 1975.
Proceedings of the 6th International Wool Textile Research Conference, South African Wool and Textile Research Institute (SAWTRI) of the Council for Scientific and Industrial Research (CSIR), Pretoria, South Africa, 1980.

Patrick T. Naughtin
Mickey G. Huson
Brett O. Bateup
Francis W. Jones
John R. Christoe
Les N. Jones
Allan G. De Boos
Keith R. Millington
Ron J. Denning
David G. Phillips
David J. Evans
Tony J. Pierlot
G. Bruce Guise
John A. Rippon
Barry V. Harrowfield
Rob A. Rottenbury
Bill Humphries
Ian M. Russell
CSIRO

WORKING FLUIDS.

See HEAT EXCHANGE TECHNOLOGY.

WORKSTATION COMPUTERS.

See COMPUTER TECHNOLOGY.

XANTHAN GUM.

See GUMS; MICROBIAL POLYSACCHARIDES.

XANTHATES

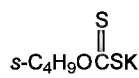
The salts of the *O*-esters of carbonodithioic acids and the corresponding *O,S*-diesters are xanthates. The free acids decompose on standing. Potassium ethyl xanthate was first prepared in 1822 by W. C. Zeise from potassium hydroxide, carbon disulfide, and ethanol. Most alcohols, including cellulose, undergo this reaction to form xanthates, but normally phenols do not (see FIBERS, REGENERATED CELLULOSICS). Potassium phenyl xanthate was prepared in 1960 from potassium phenoxide and carbon disulfide in dimethylformamide (1). The preparation of phenolic xanthates has been expanded by the use of dialkyl ethers of mono- or polyethylene glycols or sulfolane (2). Xanthates remained a laboratory curiosity until the turn of the twentieth century when the rubber industry developed a use for them in the curing and vulcanization of rubber (see RUBBER CHEMICALS; RUBBER COMPOUNDING). Cornelius Keller's invention of xanthates as flotation collectors for the nonferrous metal sulfides in 1927 (3) ranks as the chemical invention that had the greatest impact in flotation (4) (see FLOTATION). This is the principal use for the noncellulose xanthates; several of the alkali metal xanthates are commercially available.

Nomenclature

The names xanthogenic acid and xanthogen refer to the acid C₂H₅OCSSH and the radical C₂H₅OCSS[•], respectively. Xanthogenic is still used instead of xanthic and xanthogenate for xanthate, eg, sodium methyl xanthogenate [6370-03-2] for sodium methyl xanthate. The term xanthogen is sometimes used for the radical HOCSS[•], eg, methyl xanthogen acetic acid for CH₃OCSSCH₂COOH. Examples of the current nomenclature are listed in Table 1 with the corresponding common names. The latter names are used in this article.

Table 1. Nomenclature of Some Xanthic Acids and Related Compounds

Formula	CAS Registry Number	CAS and IUPAC nomenclature	Common name
$\begin{array}{c} \text{S} \\ \\ \text{CH}_3\text{O CSH} \end{array}$	[2042-42-4]	<i>O</i> -methyl carbonodithioic acid	methyl xanthic acid
$\begin{array}{c} \text{S} \\ \\ \text{C}_2\text{H}_5\text{O CSNa} \end{array}$	[140-90-9]	sodium <i>O</i> -ethyl carbono-dithioate, sodium <i>O</i> -ethyl dithiocarbonate	sodium ethyl xanthate
$\begin{array}{c} \text{S} \\ \\ \text{C}_2\text{H}_5\text{O CSCH}_3 \end{array}$	[623-54-1]	<i>O</i> -ethyl <i>S</i> -methyl carbono-dithioate	methyl ethyl xanthate
$\begin{array}{c} \text{S} \\ \\ (\text{C}_2\text{H}_5\text{O C})_2\text{S} \end{array}$	[2095-52-4]	diethyl thiodicarbonate, ((HO)C(S)) ₂ S	diethyl xanthogen monosulfide
$\begin{array}{c} \text{S} \\ \\ (\text{C}_2\text{H}_5\text{O C})_2\text{S}_2 \end{array}$	[502-55-6]	diethyl thioperoxy-dicarbonate	diethyl dixanthogen
$\begin{array}{c} \text{S} \quad \text{O} \\ \quad \\ \text{C}_2\text{H}_5\text{O CSCOC}_2\text{H}_5 \end{array}$	[3278-35-1]	diethyl thiodicarbonate, ((HO)C(O)SC(S)(OH)) 6-thioxo-3,7-dioxa-5-thio-4-nonane	diethyl xanthogen formate
$\begin{array}{c} \text{S} \\ \\ \text{C}_2\text{H}_5\text{O CNHCH}_3 \end{array}$	[817-73-2]	<i>O</i> -ethyl methylcarbamo-thioate, <i>O</i> -ethyl methyl-thiocarbamate	ethyl methylthiono-carbamate



[141-96-8]

220-260

^a Refs. 11 and 12.

When exposed to air, the sodium salts tend to take up moisture and form dihydrates. The alkali metal xanthates are soluble in water, alcohols, the lower ketones, pyridine, and acetonitrile. They are not particularly soluble in nonpolar solvents, eg, ether or ligroin. The solubilities of a number of these salts are listed in Table 4. Potassium isopropyl xanthate is soluble in acetone to ca 6 wt %, whereas the corresponding methyl, ethyl, *n*-propyl, *n*-butyl, isobutyl, isoamyl, and benzyl [2720-79-8] xanthates are soluble to more than 10 wt % (12). The solubilities of the commercially available xanthates in water are plotted versus temperature in Figure 1 (14).

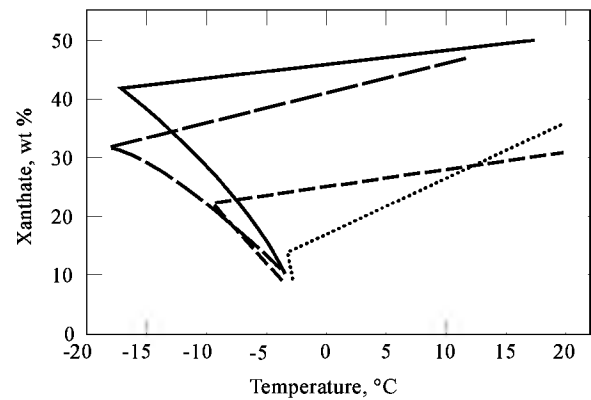


Fig. 1. Solubility of some commercial xanthates. (—), Sodium isobutyl xanthate; (), sodium ethyl xanthate; (), sodium isopropyl xanthate; (···), potassium amyl xanthate. Courtesy of Cytec Industries Inc.

Table 4. Solubilities of Some Alkali Metal Xanthates

Xanthate	CAS Registry Number	Solvent	Solubility, g 100 g soln				Refer-ence
			0°C	10°C	20°C	35°C	
sodium ethyl	[140-90-9]	water	40.8	46.0	52.0		14
potassium <i>n</i> -propyl	[2720-67-4]	water	43.0			58.0	15
		<i>n</i> -propyl alcohol	1.9			8.9	15
sodium <i>n</i> -propyl	[14394-29-7]	water	17.6			43.3	15
		<i>n</i> -propyl alcohol	10.2			22.5	15
potassium isopropyl	[140-92-1]	water	16.6			37.2	15
		isopropyl alcohol				2.0	15
		IPA-H ₂ O azeotrope			6.9		16
sodium isopropyl	[140-93-2]	water	12.1			37.9	15
			24.5	27.3	30.5		14
			24.0	27.5	31.0	37.5	17
		isopropyl alcohol				19.0	15
potassium <i>n</i> -butyl	[871-58-9]	water	32.4			47.9	15
		<i>n</i> -butyl alcohol				36.5	15
sodium <i>n</i> -butyl	[141-33-3]	water	20.0			76.2	15
		<i>n</i> -butyl alcohol				39.2	15
potassium isobutyl	[13001-46-2]	water	10.7			47.7	15
		isobutyl alcohol	1.6			6.2	15
sodium isobutyl	[25306-75-6]	water	11.2			33.4	15
			46.2	48.2	50.5		14
			44.0	49.0	51.0	57.3	16
sodium <i>sec</i> -butyl		isobutyl alcohol	1.2			20.5	15
		water	29.4	34.0	38.8		14
potassium isoamyl	[928-70-1]	water	28.4			53.5	15
			16.9	26.0	35.0		14
			28.5	39.0	45.6	52.5	16
		isoamyl alcohol	10.9			15.5	15

The heavy metal salts, in contrast to the alkali metal salts, have lower melting points and are more soluble in organic solvents, eg, methylene chloride, chloroform, tetrahydrofuran, and benzene. They are slightly soluble in water, alcohol, aliphatic hydrocarbons, and ethyl ether (18). Their thermal decompositions have been extensively studied by dta and tga (thermal gravimetric analysis) methods. They decompose to the metal sulfides and gaseous products, which are primarily carbonyl sulfide and carbon disulfide in varying ratios. In some cases, the dialkyl xanthate forms. Solvent extraction studies of a large number of elements as their xanthate salts have been reported (19).

The solubilities of the heavy metal xanthates in water have been placed in the following orders of increasing solubility by two different authors (20,21): Hg^{2+} , Ag^+ , Cu^+ , Co^{3+} , As^{3+} , Pb^{2+} , Tl^+ , Cd^{2+} , Ni^{2+} , Zn^{2+} ; and Hg_2^{2+} , Hg_2^{2+} , Au^{3+} , Ag^+ , Cu^+ , Bi^{3+} , Pb^{2+} , Cd^{2+} , Ni^{2+} , Fe^{2+} , Zn^{2+} . The solubility products of many of the salts are reported in references (9,10,22). The structure of xanthates is rather complicated. It has been shown that the formation and stability of insoluble iron xanthates depends critically on the xanthate concentration, pH, and redox potential of the solutions (23). A good literature survey and review of the chemistry of iron xanthates is given in (23). Critical micelle concentrations and Kraft points for the C_8 – C_{12} xanthates are similar to those of other ionic surfactants with equivalent chain lengths (24,25).

Alkalies stabilize xanthate solutions somewhat and the solutions readily decompose at acidic pHs (26,27). With a lower alkyl xanthate, the

decomposition rate passes through a maximum at a low pH and then decreases smoothly as the acidity increases. However, this is not the case with salts of *tert*-butyl xanthic acid [110-50-9] (28). The decomposition accelerates at pH values above 10 or below 9; between these limits, no significant decomposition occurs during 24 h (29). The kinetic studies of the rate of decomposition of alkaline xanthate solutions show a minimum rate constant at pH 10 (30). The results of these studies agree with the report that over an 8-day period, 75% decomposition occurred at pH 6.5 and only 26% at pH 10.8 (31). In Figure 2, the effect of time, temperature, and concentration on aqueous solutions of sodium isopropyl xanthate is shown (14). The stability data for the other xanthates are similar; xanthates prepared from secondary alcohols are more stable than those prepared from the primary alcohols. Branching of the chain and increasing the molecular weight also tends to increase the stability. The benzyl xanthates are less stable, both as solids and in solution.

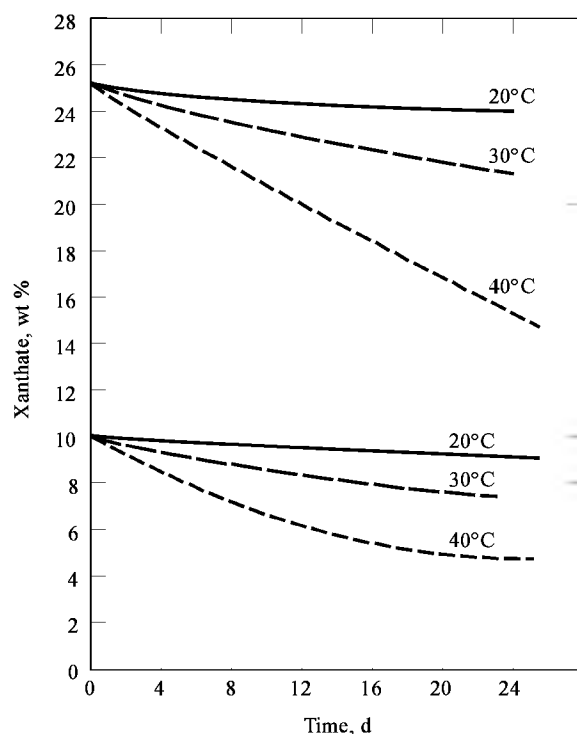
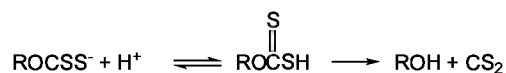


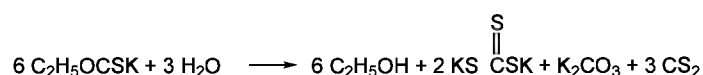
Fig. 2. Decomposition of commercial sodium isopropyl xanthate solutions. Courtesy of Cytec Industries Inc.

It was found (32) that in the acid range (pH 4–6) the alkyl group does not influence the rate of decomposition, which is similar for all xanthates. In the alkaline range the rates are markedly influenced by the substitutional group, and the rates could be correlated with the Taft polar substituent constants established for the various groups.

Reactions. The chemistry of the xanthates is essentially that of the dithio acids. The free xanthic acids readily decompose in polar solvents, the rate being 10 times greater in methanol than in hexane. The acids decompose at room temperature to carbon disulfide and the corresponding alcohol; the resulting alcohol autocatalytically facilitates the decomposition.

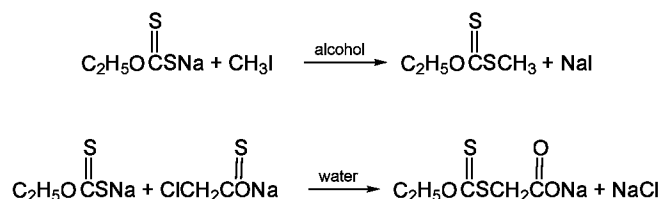


The initial hydrolysis of the xanthate in aqueous solutions at room temperature is characterized by the following reaction involving potassium ethyl xanthate:



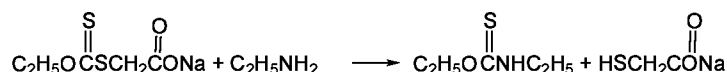
Further hydrolysis of the carbon disulfide and the trithiocarbonate produces hydrogen sulfide, etc (33). In another study of the decomposition of sodium ethyl xanthate [140-90-9] in flotation solutions, eleven components of breakdown were studied. The dependence of concentration of those components vs time was examined by solving a set of differential equations (34).

The alkali metal xanthates react readily with the various alkylating reagents to form the *S*-esters:



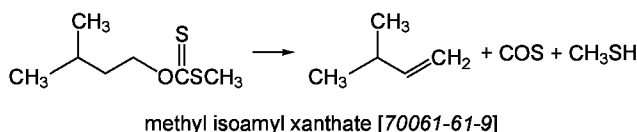
The reactions are exothermic and cooling may be required.

The dialkyl esters react readily with ammonia or alkylamines to give the corresponding thionocarbamates and thiols. In general, heating is not required. In the following reaction, the thionocarbamate is the only water-insoluble material and separates as an oil:



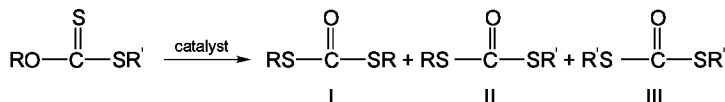
Sterically hindered amines such as *tert*-butylamine react only with this intermediate to give the desired thionocarbamate.

The Chugaev reaction, or thermal decomposition of the *S*-substituted esters of the xanthates, gives olefins without rearrangement (35,36). For example:



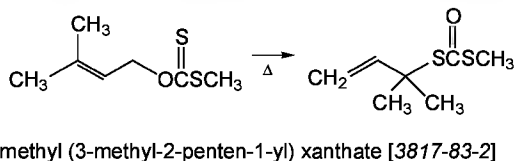
Kinetics data and Arrhenius parameters for the Chugaev reaction are given in Reference 37.

Esters derived from the primary alcohols are the most stable and those derived from the tertiary alcohols are the least stable. The decomposition temperature is lower in polar solvents, eg, dimethyl sulfoxide (DMSO), with decomposition occurring at 20°C for esters derived from the tertiary alcohols (38). Esters of benzyl xanthic acid yield stilbenes on heating, and those from neopentyl alcohols thermally rearrange to the corresponding dithiol esters (39,40). The dialkyl xanthate esters catalytically rearrange to the dithiol esters with conventional Lewis acids or trifluoroacetic acid (41,42). The esters are also catalytically rearranged to the dithioesters by pyridine *N*-oxide catalysts (43):



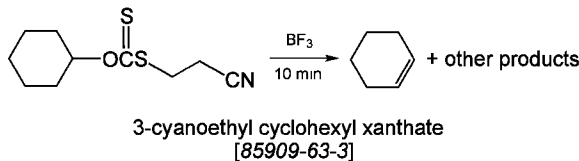
The crossover reaction products I and III should be noted (43).

Double bonds in or dialkylamino groups on the alkyl group of the *S*-methyl ester may facilitate isomerization to the dithiol ester (44). For example:

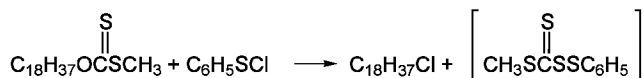


A number of catalysts of Pd(II), Pt(II), Rh(I), and Ir(I) induce rearrangements of *O*-allylic-*S*-methyl dithiocarbonates at 25°C (45).

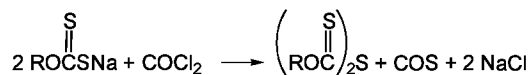
In a relatively low temperature procedure, olefins readily form from certain classes of xanthate esters (46):



The reactions of the xanthate esters with some soft electrophiles proceed with good yields (47):

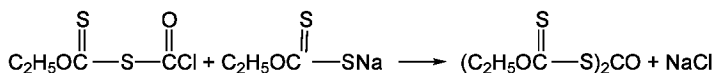
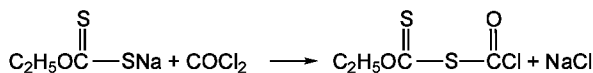


The reaction of phosgene with an aqueous solution of a xanthate and moist ether at 15–20°C gives the anhydrosulfide in good yield (48):

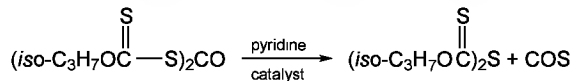


The anhydrosulfide forms in the ether layer and is readily separated.

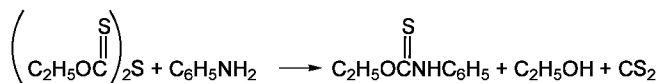
The reaction of xanthates with phosgene proceeds through the following steps (49,50):



In preparing the stable isopropyl derivative, it was found that the following rapid reaction took place (16):

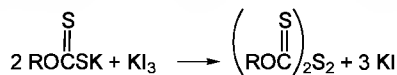


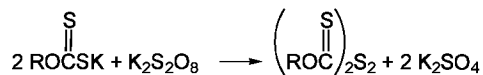
The anhydrosulfides are useful in the preparation of thionocarbamate, eg, by heating with an alcoholic solution of the arylamine:



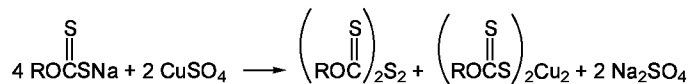
This method obviates the need to prepare the intermediate aryl isothiocyanate.

Many oxidizing agents, including sodium nitrate, convert the alkali metal xanthates to the corresponding dioxanthogen:

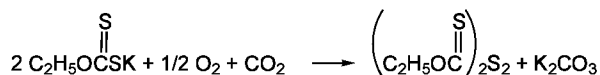




The reaction is generally carried out in water, and the resulting dioxanthogen separates as a solid or oil. Copper salts also affect the oxidation:

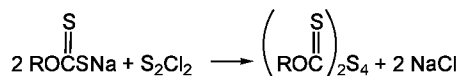


In the initial formation of the cupric xanthates, soluble xanthate complexes form prior to the precipitation of the cuprous xanthate with the concurrent formation of the dioxanthogen (51). The dioxanthogen can be separated by virtue of its solubility in ether. Older samples of alkali metal xanthates contain some dioxanthogen, which is thought to form by the following reaction (33):

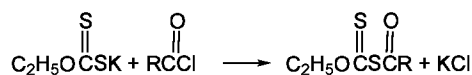


The dioxanthogens derived from the aryl xanthates are reported in Reference 52.

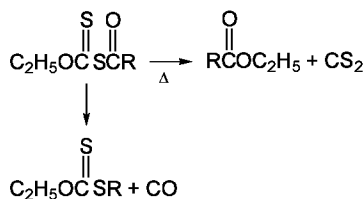
The sulfur chlorides give higher xanthate sulfides. For example, the following product is obtained from sulfur monochloride and an ether suspension at room temperature:



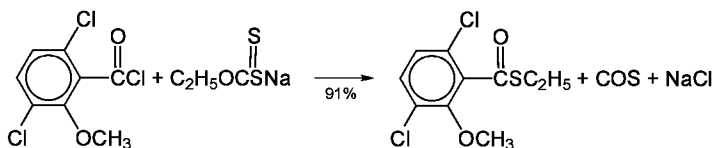
Mixed anhydrosulfides can be readily obtained by adding an acetone solution of the xanthate to an acetone solution of the acyl halide at -35°C (53):



The stability of these products varies (38). When R is methyl, the compound decomposes below room temperature, but when R is phenyl it is stable up to $40\text{--}45^\circ\text{C}$. Increasing the molecular weight of the alkyl group and introducing an electronegative group on the phenyl increases the stability. The aliphatic mixed anhydrosulfide can be decomposed by two different routes (53):

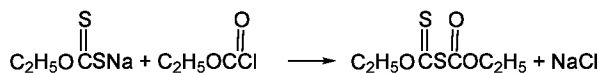


Thiol esters are obtained in 91% yield by the addition of an acyl chloride to a slurry of an alkaline salt xanthate in acetone without cooling (54):

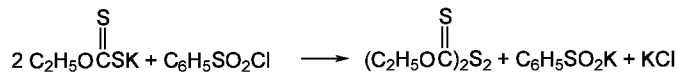


Benzoyl chloride and sodium isopropyl xanthate gave a mixed anhydride that was stable to pyridine catalyst, but a 2-year-old sample had turned to isopropyl benzoate (16).

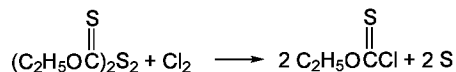
The alkyl chloroformates react with cold ethereal dispersions of the xanthates to give the fairly stable xanthogen formates.



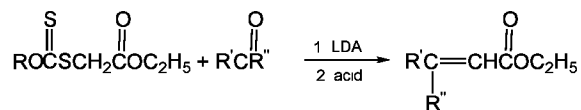
In the reaction between xanthates and sulfonyl chlorides, the xanthates convert to dioxanthogens, and the sulfonyl chlorides reduce to sulfinic acids and other compounds (38):



The acid chlorides of the xanthic acids can be prepared by the reaction of chlorine with a dioxanthogen (55):



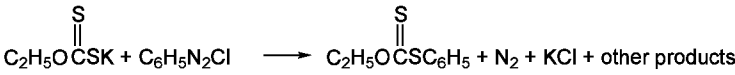
The *S*-alkoxycarbonylmethylxanthate esters treated with lithium diisopropylamide (LDA) and an aldehyde or ketone give excellent yields of α,β -unsaturated esters (56):



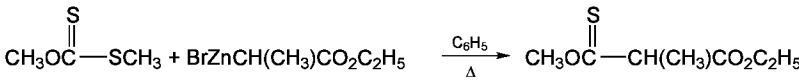
The Knoevenagel reaction is unsatisfactory for the direct preparation of these reaction products.

In the Leuckart thiophenol synthesis, the reaction of xanthates with diazonium compounds may be violent. The reaction can be controlled,

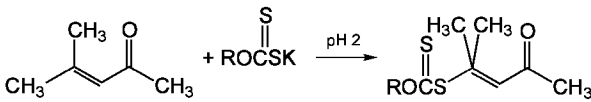
however, by thermal decomposition of the intermediate azo compounds as it forms; this provides a convenient and easy method of preparing the aryl esters, which can be saponified to the thiophenols. The crude reaction mixture is a more complex mixture than just an aryl alkyl xanthate (57):



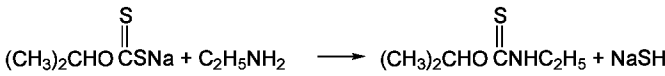
The Reformatskii reaction was carried with the dimethyl xanthate ester (58):



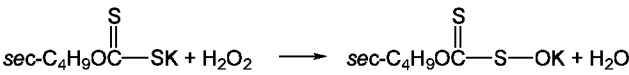
Xanthates have been added to activated double bonds, eg, acrylic derivatives and α,β -unsaturated aldehydes and ketones (59–61):



The alkali metal xanthates react with the lower alkylamines in the presence of catalytic amounts of nickel or palladium salts to give dialkylthionocarbamates (62):



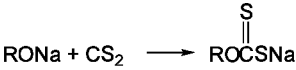
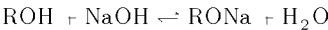
Peroxyxanthates. A new factor in the theory and practice of flotation was found in the Mount Isa, Australia, copper flotation solution. Secondary butyl perxanthate was formed by the reaction of the xanthate with hydrogen peroxide in dilute alkaline aqueous solution and was found to be identical to a substance from the flotation solution:



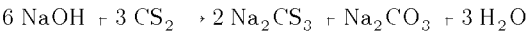
The perxanthate was isolated as the ammonium salt [69848-99-3] (63).

Preparation and Manufacture

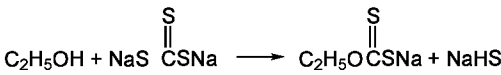
The alkali metal xanthates are generally prepared from the reaction of sodium or potassium hydroxide with an alcohol and carbon disulfide. The initial reaction is the formation of the alkoxide, which reacts with carbon disulfide to give the xanthate:



The overall heat of reaction for potassium isopropyl xanthate is 48.5 kJ mol (11.6 kcal mol) (64). The presence of water favors the principal side reaction:

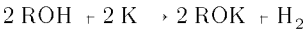


Where no water is present, very pure xanthates form (65). Ethanol reacts slowly with sodium trithiocarbonate to produce sodium ethyl xanthate (66):



The nature of the inorganic by-products in technical xanthates varies, depending upon their exposure to air. Thus, Na₂S, Na₂SO₄, and Na₂S₂O₃ may be present in older samples exposed to air (67).

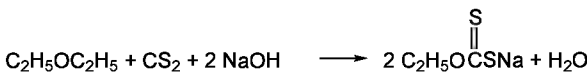
In the case of tertiary and some of the more complex alcohols, the use of alkali hydroxides is not feasible, and it is necessary to use reagents such as sodium hydride, sodium amide, or the alkali metal to form the alkoxide:



Powdered KOH in DMF (dimethylformamide) or DMSO reacts with carbon disulfide to give these xanthates.

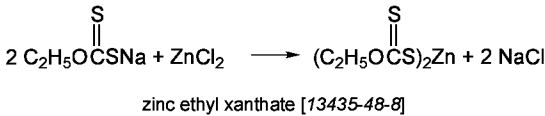
Various xanthates can be readily prepared in the laboratory from alcohols (up to C₁), potassium hydroxide, and carbon disulfide (68).

Another, quite different procedure is the preparation of xanthates from ethers (69):



In a study of the influence of temperature (30–45°C) on the preparation of isopropyl xanthates, it was determined that increasing the temperature resulted in a decrease in the xanthate yield and an increase in by-products. Also, a decrease in the water content of the alcohol increases the xanthate yield (70).

Many of the heavy metal xanthates have been prepared from aqueous solutions of the alkali metal xanthates and the water-soluble compound of the heavy metal desired.



A modified and detailed procedure is given for the preparation of the corresponding Cr³⁺, In³⁺, and Co³⁺ xanthates in good yields and high purity (18). A structural study review of inorganic xanthates (29 metals) of the common xanthates was presented. The emphasis was on single-crystal x-ray methods; some solid-state nmr studies were included. The average C–S bond distance of 1.682 × 10^{–4} microns is intermediate between the values expected for a single bond distance of 1.81 × 10^{–4} microns and a C=S distance of 1.62 × 10^{–4} microns (71).

Alkali Metal Xanthates. The commercially available xanthates are prepared from various primary or secondary alcohols. The alkyl group varies from C₂ to C₈ and the alkali metal may be sodium or potassium. Not all of the commercially available alcohols in the C₂–C₈ range are available as their xanthates, but most could be made if there were sufficient demand for them. Except for a few foreign articles and the old reports of the U.S. Office of Technical Services, most information on the manufacture of xanthates is in the patent literature. The important patent on flotation uses of xanthates appeared in 1927 (3), and all articles and patents on the manufacture of xanthates have appeared since then (72–82).

Steel (qv) generally is suitable as a material of construction for the apparatus used in the manufacture of xanthates. Cooling is required to minimize side reactions. The reaction temperature is generally maintained below 40°C. In one procedure, the reaction is carried out in an inert diluent, eg, petroleum ether (72–76). A plant in the former Federal Republic of Germany employed this method for the preparation of sodium amyl and sodium hexyl xanthate. The copper kettles were equipped with stirring and cooling apparatus. A petroleum ether slurry was made of equimolar quantities of powdered sodium hydroxide and the alcohol. Carbon disulfide was added and the temperature was maintained at ca 35–40°C. After 3 h, the product was separated by filtration, washed with petroleum ether, and dried. The Dow Chemical xanthate plant, which is no longer in operation, in a similar manner used pentane as a diluent and the heat of reaction was dissipated by the refluxing of the pentane (bp 36°C), which prevented overheating of the reaction mixture. The reactants and diluent were added to four steel reactors in series, which were equipped with stirrers and condensers. The product was fed from the last reactor to a spray dryer, then to a pelletizer, and finally to the drumming equipment. The spray dryer minimized the thermal decomposition of the product in the drying step. The volatile organic materials were recycled. The annual capacity of the plant was 13,600 metric tons of xanthates.

For the manufacturing of potassium ethyl xanthate, 400% excess of alcohol and equimolar quantities of 50 wt % aqueous potassium hydroxide and carbon disulfide were used (77). After 30 min at 40°C, the mixture was vacuum drum dried. The product was obtained in near quantitative yield and assayed at 95%. It is claimed that potassium amyl xanthate can be made with almost the same ratio of reactants and 80 wt % caustic potash (78).

Xanthates have been synthesized from C₂–C₈ oxo alcohols in a reactor with intensive stirring at 25°C. The water and unreacted alcohol are removed from the product at 40–50°C under vacuum to give a friable powder of 80% purity in 77% yield (79) (see OXO PROCESS; ALCOHOLS, HIGHER ALIPHATIC).

In another process, 50–150% excess carbon disulfide and a small excess of powdered alkali hydroxide are added, with stirring and cooling to the lower alkyl alcohol. After completion of the reaction, the excess CS₂ and resulting water of reaction are removed by applying a vacuum to the reactor (80).

The most important hazard in the manufacturing of xanthates is the use of carbon disulfide (qv) because of its low flash point, ignition temperature, and its toxicity. A report on the manufacture of sodium ethyl xanthate at Kennecott Nevada Mines Division discusses the various safety problems and the design of a facility (81). A plant layout and a description of the reagent preparations are also given.

One patent describes a continuous process involving an aqueous alkali metal hydroxide, carbon disulfide, and an alcohol (82). The reported reaction time is 0.5–10 min before the mixture is fed to the dryer. The usual residence time is on the order of hours. A study in the former USSR reported the use of the water–alcohol azeotrope for water removal from isobutyl or isoamyl alcohol and the appropriate alkali hydroxide to form the alkoxide prior to the addition of carbon disulfide (83).

Because of hydrate formation, the sodium salts tend to be difficult to dry. Excess water over that of hydration is believed to accelerate the decomposition of the xanthate salts. The effect of heat on the drying of sodium ethyl xanthate at 50°C has been studied (84):

Time of heating, h	Xanthate content, wt %
0	79.32
40	74.38
64	65.66
88	65.06
112	63.85
136	60.41

Economic Aspects

The two main producers of xanthates in the United States, ie, American Cyanamid and Dow, shut down their plants in the 1970s. American Cyanamid transferred its production to Canada, and since then has shut down that plant. Xanthates in solution are available in some areas of the United States and Canada and are transported locally by truck. The U.S. consumption of xanthates for froth flotation was ca 1,812 t yr in 1985 (85). In the previously non-Soviet bloc the 1995 xanthate usage was estimated at 30,000 metric tons, and in China and Russia at 25,000 metric tons, for a worldwide total of 55,000 metric tons. The estimated current number of xanthate plants in the world is 14–16, with one in the United States and one in Canada.

General price information for 1996 on commercially available xanthates from a Canadian company is given in Table 5. Prices are approximate based on truckload shipments within the continental United States (excluding Alaska).

Table 5. Prices of Commercially Produced Canadian Xanthates, 1996^a

Xanthate	Price, \$ /kg
sodium isopropyl	1.50
potassium ethyl	1.32
potassium isobutyl	1.58
potassium amyl	1.54

^a Both the sodium and potassium salts are available in drums, FIBCs and boxed FIBCs.

Standards

The alkali metal xanthates are obtainable in the technical grade only, and they vary in particle size from powder to pellets. The color is any of various shades of yellow and the odor is unpleasant. There are no published specifications for the xanthates, but each manufacturer has its own standards. The assay on the commercial materials is usually ca 90–95%. The assay does not necessarily indicate the collector ability of a xanthate (see FLOTATION). Although all xanthates decompose to some extent on prolonged storage, such decomposition does not necessarily result in a corresponding reduction in mineral-collecting power. This lack of correlation between xanthate content and flotation-collecting ability has been noted in products produced by different methods. Apparently, some of the xanthate oxidation or decomposition products are effective as flotation agents.

The material is shipped in steel drums. The U.S. Department of Transportation (DOT) regulations for the xanthates is the hazardous material

description ORM—A (other regulated material—regulated by air), NOS.

Analytical Methods

Although there are several published analytical procedures, many are either tedious or inaccurate. For example, when assayed by the standard KI₃ method, the copper sulfate method, or the aqueous acid method, certain of the inorganic impurities assay as xanthate, thereby giving high results (86,87). When specifications are drawn up for xanthate content, the method of assay should be given. In a satisfactory procedure developed by The Dow Chemical Company, the xanthate content is determined by dissolving the sample in acetone, removing the inorganic solids by filtration, decomposing the xanthate with an excess of standard acid, and back-titrating with a standard base. The components of the inorganic solid left on the filter can be determined by any standard procedure. The water content can be determined by the Karl Fischer method (88). The nonxanthate organic material can be determined by extraction with ether. The uv absorption spectra of xanthates have been extensively used for the assay of dilute solutions of the latter (26,89). The interference of trithiocarbonates can be avoided through the formation of a nickel xanthate. The latter is extracted into heptane and the absorbance spectrum (425 nm) is measured (90).

The successful separation of xanthate-related compounds by high performance liquid chromatography (hplc) methods has been reported (91–93). The thin-layer chromatography procedure has been used to determine the nature of the alcohols in a xanthate mixture. A short run of 3 cm at a development time of 25 min gives a complete separation of C₁–C₅ alkanol xanthates (94).

The analysis of solutions of technical xanthates by Ag⁺ potentiometric titration, with the addition of ammonium hydroxide, has been successfully used at Dow (95).

The separation of xanthates by ion-interaction reversed-phase column chromatography is described for the determination of eight different xanthates in reagents commonly used in flotation plants. The separated species were detected spectroscopically at a wavelength of 305 nm (96).

Health and Safety Factors, Toxicology

Most of the published data on the toxicity of the xanthates are given in the foreign literature. The xanthates are low in acute and oral toxicity, as indicated in Table 6. Potassium amyl xanthate causes extensive pain and slight corneal injury to the eye and may burn the skin on prolonged contact. In a chronic toxicity study of this xanthate, an aqueous aerosol was applied to dogs, rabbits, rats, and mice. There were no adverse effects in the latter three at 23 mg m³, but liver damage occurred in dogs exposed to this level. A no-effect level was not determined. The recommended airborne industrial hygiene guide (IHG) is 1 mg m³ (100). The maximum permissible concentrations recommended for ethyl xanthogenate is 0.5 mg m³ and 1.0 mg m³ for isopropyl, isobutyl, and isoamyl xanthogenates (102). Note: Consumption of alcohol and exposure to xanthates have a synergistic effect (103).

Table 6. Oral Toxicity of Xanthates

Xanthate	Species	LD ₀ , mg kg	LD ₅₀ , mg kg	References
sodium ethyl	rat	500		97
potassium ethyl	rat	500	1700	97,98
	mouse		583	98
sodium isopropyl	rat	250		97
potassium isopropyl	rat		1700	98
	mouse		583	98
potassium <i>n</i> -butyl	mouse		411, 465	99,100
sodium isobutyl	rat	500		97
potassium isobutyl	rat		1290	98
	mouse		480	98
sodium <i>sec</i> -butyl	rat		>2000	97
potassium amyl (mixed)	rat	1000	1000–2000	97,101
potassium isoamyl	rat		765	98
	mouse		470	98

The alkali metal xanthates are fairly safe to handle. The standard precautions of rubber gloves, dust mask, and goggles are sufficient when handling the solid or the solution. If xanthates come in contact with the body, dermatitis may result, although susceptibility varies from person to person. Contaminated clothing should be laundered before being worn again. Internally, whether in the form of dust or fumes, the xanthates act similarly to carbon disulfide. The possible formation of carbon disulfide in xanthate solutions requires that proper care be taken in the preparation and handling of these solutions to avoid possible fires and explosions (14).

Under regulations for the enforcement of the Federal Insecticide, Fungicide, and Rodenticide Act, products containing over 50 wt % sodium isopropyl xanthate must bear the label "Caution. Irritating dust. Avoid breathing dust, avoid contact with skin and eyes" (104). Rubber goods in repeated contact with food may contain diethyl xanthogen disulfide not to exceed 5 wt % of the rubber products. The upper limit for butyl dixanthogen [105-77-1], for diisopropyl xanthogen polysulfide (a 1:2:1 mixture of the trisulfide [52584-27-7], tetrasulfide [69303-50-0], and the pentasulfide), and for zinc butyl xanthate [150-88-9] is 1.5 wt % (105,106).

Xanthate drums should be kept as cool and dry as possible. Protection from moisture is the most important factor. A combination of moisture and hot weather causes sodium ethyl xanthate to ignite spontaneously (14).

Environmental Concerns. Concern for the well-being of the environment has resulted in studies on the effects of mining chemicals, including the xanthates, on various aquatic organisms worldwide (107–110). In a thorough and detailed study of three typical mill operations, it was concluded that residual organic flotation reagents do not seem to present widespread problems in effluent disposal. A large portion of the reagents used never reached the final discharge. Biodegradation appears to be the commonest method of reagent removal. At levels below 20–25 mg L, the xanthates are biodegradable (111).

Uses

Besides the importance of cellulose xanthates in the manufacture of rayon and cellophane, the primary use for the alkali metal xanthates is as collectors in the flotation of metallic sulfide ores. The xanthates are the principal metal sulfide collectors. A great deal of experimental work on the production methods and potential uses of the starch xanthates has been reported (112). Allyl amyl xanthates are still useful reagents in the flotation of copper–molybdenum sulfide ores. Other uses of the xanthates are very minor. This is not so much the result of the lack of activity or performance of these compounds as it is the fact that superior products are available. Xanthates and derivatives have been recommended for the vulcanization of rubber (see RUBBER CHEMICALS; RUBBER COMPOUNDING), as herbicides (qv), insecticides, fungicides, high pressure lubricant additives, and for analytical procedures (see INSECT CONTROL TECHNOLOGY; FUNGICIDES, AGRICULTURAL; LUBRICATION AND LUBRICANTS).

In addition to a continued increase in the number of use patents in these fields, a new use of xanthates as inhibitors of fertilizer nitrogen transformation in soil has been reported, as well as the use of certain metal xanthates as color developers for image-recording materials (113,114) (see FERTILIZERS; COLOR PHOTOGRAPHY). For several years, sodium isopropyl xanthate was used as an intermediate in the manufacture of saccharin (see

SWEETENERS).

Xanthates are used in a froth flotation process of soils contaminated with mercury. The soil to be treated is run through hydrocyclones, and the slurries are flocculated, dewatered, and removed to a secure landfill. The effluent water is recycled. The process is suitable for treating industrial land sites contaminated with mercury droplets (115).

Derivatives

The principal derivatives are of four types. The dixanthogens, shown herein, were



produced at one time as additives for polymers and isopropyl dixanthogen is now available from a Mexican producer of xanthates. The mixed anhydrosulfides, ie, the xanthogen formates, are used in acid pulp flotation of sulfide ores. The most important one, diethyl xanthogen formate, has been used in Chile for over 50 yr. Dialkyl thionocarbamates are used in the flotation of copper and zinc sulfides, in particular where the rejection of iron pyrites is important. The two principal products are isopropyl ethylthionocarbamate and butyl ethylthionocarbamate. The *S*-alkyl xanthate esters (ROCS—S—CH₂—CH=C₄) have been found to be particularly effective on molybdenite and copper–molybdenite ores.

BIBLIOGRAPHY

"Xanthic Acids and Xanthates" in *ECT* 1st ed., Vol. 15, pp. 150–157, by G. H. Harris, The Dow Chemical Company; "Xanthates" in *ECT* 2nd ed., Vol. 22, pp. 419–429, by G. H. Harris, The Dow Chemical Company; in *ECT* 3rd ed., Vol. 24, pp. 645–661, by G. H. Harris, Consultant.

1. A. F. McKay, D. L. Garmaise, G. Y. Paris, S. Gelblum, and R. J. Ranz, *Can. J. Chem.* **38**, 2042 (1960).
2. **U.S. Pat. 5,239,104** (Aug. 24, 1993), J. E. Shaw (to Phillips Petroleum Co.).
3. **U.S. Pat. 1,628,151** (May 10, 1927), C. Keller.
4. D. W. Fuerstenau, *Proceedings 19th Int. Min. Proc. Cong., 1995*, Vol. 3, San Francisco, 1995, pp. 3–17.
5. I. Iwasaki and S. R. H. Cooke, *J. Phys. Chem.* **63**, 1321 (1959).
6. V. Hejl and F. Pechar, *Chem. Zvesti* **21**, 261 (1967).
7. A. P. Gavrish, *Ukr. Khim. Zh.* **29**, 900 (1963).
8. A. P. Sanzharova, *Ukr. Khim. Zh.* **35**, 91 (1969).
9. K. Hayashi, Y. Sasaki, S. Inomato, and Y. Takagami, *Bull. Chem. Soc. Jpn.* **57**, 3074–3076 (1984).
10. S. Doi, J. Shibata, and M. Sano, *Technol. Rep. Kansai Univ.* **28**, 105–114 (1986).
11. I. Tyden, *Talanta* **13**, 1353 (1966).
12. I. S. Shupe, *J. Assoc. Offic. Agric. Chem.* **25**, 495 (1942).
13. K. G. Rutherford, R. M. Ottenbrite, and B. K. Tang, *J. Chem. Soc. (C)*, 582 (1971).
14. *Xanthate Handbook*, American Cyanamid Company, Mining Chemicals Department, Wayne, N.J., 1972.
15. I. Yu. Keskyula, S. B. Faerman, Ch. I. Kondrat'ev, E. L. Goncharova, R. M. Sorokina, and K. K. Chevychalova, *Trans. State Inst. Appl. Chem. (USSR)* **30**, 68 (1936).
16. Technical data, G. H. Harris, The Dow Chemical Co.
17. M. W. Clark and L. O. Lui, personal communication, 1969.
18. F. Galsbol and C. E. Schaffer, *Inorg. Synth.* **10**, 42 (1967).
19. E. M. Donaldson, *Talanta* **23**, 411, 417 (1976).
20. L. Malatesta, *Chim. Ind. (Milan)* **23**, 319 (1941).
21. A. T. Pilipenko, *Zh. Anal. Khim.* **4**, 227 (1949).
22. I. A. Kakovskii, *Proceedings of the International Congress on Surface Activity, 2nd, London*, Vol. 4, 1957, pp. 225.
23. X. Wang, K. S. Eric Frossberg, and N. J. Bohin, *Int. J. Miner. Processing* **27**, 1–19 (1989).
24. C. C. DeWitt, R. F. Makens, and A. W. Helz, *J. Am. Chem. Soc.* **57**, 796 (1935).
25. I. C. Hamilton and R. Woods, *Int. J. Miner. Processing* **17**, 113–120 (1986).
26. J. Dyer and L. H. Phifer, *Macromolecules*, 111 (1969).
27. R. J. Millican and C. K. Sauers, *J. Org. Chem.* **44**, 1964 (1979).
28. C. A. Bunton, P. Ng, and L. Sepulvada, *J. Org. Chem.* **39**, 1130 (1974).
29. P. J. C. Fierens, J. Adam, and E. Royers, *Ind. Chim. Belge Suppl. 1959* **2**, 777 (1959).
30. N. P. Finkelstein, *Colloid Polym. Sci.* **255**, 168 (1977).
31. I. Iwasaki and S. R. B. Cooke, *J. Am. Chem. Soc.* **80**, 285 (1958).
32. P. J. Harris, *S. Afr. J. Chem.*, 37, 91–95 (1984).
33. A. Pomianowski and J. Leja, *Can. J. Chem.* **41**, 2219 (1963).
34. H. Dautzenberg, A. Kowal, J. Kowal, and W. Deitzel, *Reagents Miner. Ind. Pap.*, 47–51 (1984).
35. G. L. O'Connor and H. R. Nace, *J. Am. Chem. Soc.* **74**, 5454 (1952); **75**, 2118 (1953).
36. H. R. Nace in A. C. Cope, ed., *Organic Reactions*, Vol. 12, John Wiley & Sons, Inc., New York, 1962.
37. A. I. Awadi, B. Nouria, and B. David, *J. Chem. Soc. Perkin Trans.* **2**, 773–775 (1982).
38. P. Meurding, K. Sjoeborg, and B. Sjoeborg, *Acta Chim. Scand.* **26**, 279 (1972).
39. G. Bulmer and F. G. Mann, *J. Chem. Soc.*, 666, 677, 680 (1945).
40. K. G. Rutherford, B. K. Tang, L. K. M. Lam, and D. P. C. Fung, *Can. J. Chem.* **50**, 3288 (1972).
41. K. Komaki, T. Kawata, K. Harano, and T. Taguchi, *Chem. Pharm. Bull.* **26**, 3807 (1978).
42. M. W. Fichtner and N. F. Haley, *J. Org. Chem.* **46**, 3141 (1981).
43. K. Horano, H. Nakagawa, K. Kamei, H. Kiyonega, and T. Hisano, *Chem. Pharm. Bull.* **40**, 1675–1682 (1992).
44. T. Taguchi and co-workers, *Tetrahedron Lett.*, 2717 (1965).
45. P. R. Auburn, J. Whelan, and B. Bosnich, *Isr. J. Chem.* **27**, 250–254 (1986).
46. *Chem. Eng. News* **47**, 41 (Sept. 22, 1969).
47. D. H. R. Barton, R. V. Stick, and R. Subramanian, *J. Chem. Soc. Perkin Trans. 1*, 2112 (1976).
48. S. V. Zhuravlev and M. I. Dalchenko, *Zhur. Priklad. Khim. (J. Applied Chem. USSR)* **20**, 1038–1043 (1947).
49. Sh. Kh. Khalikov, S. V. Alieva, and M. M. Sobirov, *Khim. Priir. Soedin.*, 844–845 (1985).
50. **U.S. Pat. 1,556,115** (Apr. 15, 1985), J. E. Franz and K. E. Zwickelmaier.
51. G. Sparrow, A. Pomianowski, and J. Leja, *Sep. Sci.* **12**, 87 (1977).
52. H. W. Chen, J. P. Fackler, Jr., D. P. Schussler, and L. D. Thompson, *J. Am. Chem. Soc.* **100**, 2370 (1978).
53. D. H. R. Barton, M. V. George, and M. Tomoeda, *J. Chem. Soc.*, 1967 (1962).
54. J. J. D'Amico, T. Shafer, and D. Jacobson, *Phosph. Sulfur* **7**, 301 (1979).

55. D. Martin and W. Mucke, *Chem. Ber.* **98**, 2059 (1965).

56. K. Tanaka, R. Tanikaga, and A. Kaji, *Bull. Chem. Soc. Jpn.* **52**, 3619 (1979).

57. J. R. Cox, Jr., C. L. Gladys, L. Field, and D. E. Pearson, *J. Org. Chem.* **25**, 1083 (1960).

58. M. Chandrasekharan, L. Bhat, I. Hiriyakkanavar, and J. Hiriyakkanavar, *Tetrahedron Lett.* **34**, 6439–6442 (1993).

59. O. Bayer, *Angew. Chem.* **61**, 229 (1949).

60. J. L. Garraway, *J. Chem. Soc.*, 4072 (1962).

61. S. V. Tsarenko, P. P. Gnatyuk, V. A. Ignatov, and V. A. Malii, *Izv. Vyssh. Uchebn. Zaved. Khim. Khim. Tekhnol.* **23**, 836 (1980).

62. **U.S. Pat. 3,975,264** (Aug. 17, 1976), F. A. Booth, R. D. Crozier, and L. E. Strow (to Minerec Corp.).

63. M. H. Jones and J. T. Woodcock, *Int. J. Miner. Process.* **5**, 285–296 (1978).

64. V. P. Savel'yanov, I. P. Utrobin, R. T. Savel'yanova, and N. N. Rvseev, *Russ. J. Phys. Chem.* **54**, 720 (1980).

65. **U.S. Pat. 2,024,923** (Dec. 17, 1935), W. Hirschkind (to Great Western Electro-Chemical Co.).

66. G. Ingram and B. A. Toms, *J. Chem. Soc.*, 4328 (1957).

67. S. N. Danilov, N. M. Grad, and E. I. Geine, *Zh. Obshch. Khim.* **19**, 826 (1949).

68. L. S. Foster, *Technical Paper No. 2*, Department of Mining and Metallurgy Research, University of Utah, Salt Lake City, Utah, 1928.

69. **U.S. Pat. 2,534,085** (Dec. 12, 1950), B. M. Vanderbilt and J. P. Thorn (to Standard Oil Development Co.).

70. T. T. Savel'yanova, I. P. Utrobin, N. N. Zaitseva, and U. P. Savel'yanova, *J. Appl. Chem. USSR* **53**, 1855 (1980).

71. E. R. T. Tiekink and G. Winter, *Rev. Inorg. Chem.* **12**, 183–302 (1993).

72. W. Hensinger, *U.S. Office of Technical Services Report PB 74736*, U.S. Department of Commerce, Washington, D.C., 1934–1936.

73. **U.S. Pat. 1,559,504** (Oct. 27, 1925), R. B. Crowell and G. F. Breckenridge (to Western Industries Co.).

74. **U.S. Pat. 2,107,065** (Feb. 1, 1938), A. J. Van Peski (to Western Industries Co.).

75. **Ger. Pat. 1,934,175** (Jan. 14, 1971), K. Baessler and G. Folz (to Farberwerke Hoechst AG).

76. **USSR Pat. 727,641** (Apr. 15, 1980), V. P. Savel'yanova, R. T. Savel'yanova, N. P. Utrobin, and T. V. Rudakova (to D. I. Mendeleev Chem-Tech, Institute).

77. W. A. M. Edwards and J. H. Clayton, *U.S. Office of Technical Services Report PB 34028*, U.S. Department of Commerce, Washington, D.C., 1945, p. 28.

78. **Ger. Pat. 765,196** (July 28, 1955), H. Gunther (to Deutsche Gold- und Bilker-Scheideanstalt vorm Roessler).

79. G. A. Yefimova, Ye. I. Silina, and V. I. Ryaboy, *Sov. Chem. Ind.* **47**, 550 (1971).

80. **Ger. Offen. 2,149,726** (Apr. 6, 1972), G. J. Novak and A. J. Robertson (to American Cyanamid Co.).

81. J. J. Harrington and P. B. Valenti, *1976 Fall Meeting, American Institute of Mining Engineers Paper No. 76B-358*, Engineering Societies Library, New York, 1976.

82. **Czech. Pat. 97,929** (Jan. 15, 1961), O. Leminger and co-workers,

83. N. I. Gel'perin and co-workers, *Sb. Nauchn. Tr. Gos. Nauchn. Issled. Inst. Tsvetn. Met.* 170 (1959).

84. **U.S. Pat. 1,724,549** (Aug. 13, 1929), T. W. Bartram and W. C. Weltman (to Rubber Service Laboratories Co.).

85. *Minerals Yearbook*, Vol. 1, U.S. Department of the Interior, Bureau of Mines, U.S. Government Printing Office, Washington, D.C., 1985, p. 42.

86. W. S. Calcott, F. L. English, and F. B. Downing, *Eng. Min. J. Press* **118**, 980 (1924).

87. W. Hirschkind, *End. Min. J. Press* **119**, 968 (1925).

88. A. L. Linch, *Anal. Chem.* **23**, 293 (1951).

89. R. Joedodibroto, Ph.D. dissertation, Syracuse University, Syracuse, N.Y., 1963.

90. D. Kyriacou, personal communication, 1969.

91. J. G. Eckhardt, K. Stetzenbach, M. F. Burke, and J. L. Moyers, *J. Chromatogr. Sci.* **16**, 510 (1978).

92. R. T. Honeyman and R. R. Schrieke, *Anal. Chim. Acta* **116**, 345 (1980).

93. C. Zhou, H. Cai, and Y. Li, *Fresenius J. Anal. Chem.* **338**, 908–911 (1990).

94. A. Messina and D. Corradini, *J. Chromatogr.* **207**, 152 (1981).

95. H. O. Kedinger, personal communication, 1982.

96. E. D. Barnes and C. Pohlarde-Watson, *Rep. MINTEK*, M367, 12 pp. (1988).

97. Technical data, The Dow Chemical Company, C. B. Shaffer, American Cyanamid, 1964; P. Avotin, personal communication, 1982.

98. E. M. Trofinovich, S. M. Rykova, M. A. Molchanova, and L. B. Aleksandrovskaia, *Gig. Sanit.* (6), 95 (1976).

99. E. A. Babayan, *Materialy 2-oi[Vtoroi] Itog. Nauchn. Konf. Inst. Gigieny Truda i Prof. Zabolevan. Posryashch. Vopr. Gigieny Truda i Prof. Patol. Erevan, Sb.* 1963, 75 (1964).

100. E. A. Babayan, *Mater. Itogovoi Nauch. Konf. Vop. Gig. Tr. Profpatol. Khim. Gornorud. Prom., 3d 1966*, 97 (1968).

101. N. G. Fronk, personal communication, 1982.

102. R. D. Ovanesyan, *Zh. Eksp. Klin. Med.* **24**(4), 344–348 (USSR) (1984).

103. T. Robinson and H. Jones, *Publ. Austral. Inst. Min. Metall.*, **9/92** (*Extraction Metallurgy of Gold and Base Metals*), 121–127 (1992).

104. *Fed. Regist.* **27**, 2267 (Mar. 9, 1962).

105. 21 CFR Ch. 1, *HHS Publication*, U.S. FDA, Apr. 1, 1996, §177.2600.

106. *Fed. Regist.* **3**, 39647-8 (Aug. 1995).

107. M. C. Fuerstenau, *Trans. Soc. Min. Eng. AIME* **255**, 337 (1974).

108. G. Leduc and co-workers, *Arctic Land Use Research Report, ALUR, 74-75-34*, Ottawa, Canada, 1976.

109. A. Solak, A. Wernikowska-Ukleja, and H. Multan, *Cuprum* **6**(2), 32 (1979).

110. M. V. Mosevich and co-workers, *Izv. Gos. Nauchn. Issled. Inst. Ozhern. Rechn. Rybn. Khoz.* **109**, 80 (1976).

111. A. D. Read and R. M. Manser, *Proc. Int. Miner. Process. Cong., 11th*, 1323 (1975).

112. Numerous patents and publications of the Northern Regional Research Center, Peoria, Ill.

113. J. Ashworth, G. A. Rodgers, and G. G. Briggs, *Chem. Ind. (London)*, 90 (1979).

114. **Jpn. Kokai Tokyo Koho 79 10,009** (Jan. 25, 1979), M. Satomura and K. Sato (to Fuji Photo Film Co., Ltd.).

115. **U.S. Pat. 5,244,492** (Sept. 14, 1993) Benoit Cyr (to PPG Industries Inc.).

General References

Reference 14 is also a general reference.

G. Gattow and W. R. Bahrendt, *Topics in Sulfur Chemistry, Vol. 2: Carbon Sulfides and Their Inorganic and Complex Chemistry*, Georg Thieme, Stuttgart, Germany, 1977.

G. H. Harris, *Reagents in Mineral Technology*, Surfactant Science Series, Vol. 27, Marcel Dekker, Inc., New York, 1988, pp. 371–383.

J. Leja, *Surface Chemistry of Froth Flotation*, Plenum Press, New York, 1982.

S. R. Rao, *Xanthates and Related Compounds*, Marcel Dekker, Inc., New York, 1971.

Guy H. Harris
University of California at Berkeley

XENON.

See HELIUM GROUP, GASES.

XENON COMPOUNDS.

See HELIUM GROUP, GASES.

XEROGRAPHY.

See ELECTROPHOTOGRAPHY.

X-RAY ANALYSIS.

See X-RAY TECHNOLOGY.

X-RAY TECHNOLOGY

Analytical x-ray instruments are used to characterize materials in several different ways. As with medical x-ray instruments there are analytical instruments that can produce images of internal structures of objects that are opaque to visible light. There are instruments that can determine the chemical elemental composition of an object, that can identify the crystalline phases of a mixture of solids, and others that determine the complete atomic and molecular structure of a single crystal. These are the most common applications for x-ray instruments.

The determination of particle size and structural information for fibers and polymers, and the study of stress, texture, and thin films are applications that are growing in importance and can be examined with x-ray instruments.

Because of recent advances in hardware, particularly detector hardware, and computer software, x-ray instruments have become very powerful. Problems that could not be solved several years ago can be solved with the newer instrumentation. Also the instruments have become much more automatic so that the more routine problems can be solved much faster than a few years ago.

Characterization and Generation of X-Rays

X-Ray Electromagnetic Spectrum. X-rays are a form of electromagnetic radiation and have a wavelength, λ , much shorter than visible light. The center of the visible light spectrum has a wavelength of about 0.56×10^{-6} m. The range of wavelengths for x-rays used in the applications discussed here are from about 0.01×10^{-9} m to about 7.0×10^{-9} m. The energy of an x-ray photon expressed in kiloelectron-volts ($1.0 \text{ keV} = 1000 \text{ eV}$) is related to wavelength expressed in nanometers ($1.0 \text{ nm} = 1.0 \times 10^{-9}$ m) as follows:

$$E(\text{keV}) \approx 1.24 / \lambda(\text{nm})$$

There are several ways to produce x-rays that have the range of energies appropriate for the various applications discussed here. The most commonly used methods are discussed below.

Synchrotron Radiation. X-rays are produced when very energetic electrons traveling close to the speed of light are decelerated. A continuous x-ray spectrum of intensity vs wavelength is produced. Very intense x-ray sources are available at synchrotron installations around the world. The Stanford Linear Accelerator in California is typical of these installations. Electrons are accelerated with electromagnets while traveling along a linear path of 2 miles. Then they are inserted into a nearly circular path which is maintained by bending magnets. In this circular path, the electrons lose energy by producing x-ray photons whose paths are tangential to the circle. X-ray instruments designed for many different applications are placed around the synchrotron ring.

Synchrotrons were originally designed for high energy particle physics studies. The synchrotron x-ray radiation was not desirable for those studies. Later synchrotrons were designed not for high energy particle physics studies, but instead were designed to optimize the desirable features of the x-ray and ultraviolet radiation produced. The so-called third-generation synchrotrons are the European Synchrotron Radiation Facility (ESRF) in Grenoble, France, The Advanced Photon Source (APS) at Argonne in the United States and the Super Photon Ring (Spring-8) in Harima, Japan. These facilities produce high energy x-ray beams at least 10,000 times brighter than the early synchrotrons. They have straight sections in their otherwise circular paths to allow the insertion of wiggler and undulator magnets. These magnets cause the electrons to oscillate in the plane of the ring causing a large increase in the intensity of the synchrotron radiation.

An undulator produces x-ray radiation as sharp peaks at several harmonics of some fundamental radiation energy. At higher magnetic field strength, the higher harmonics tend to smooth out producing a continuous and very intense spectrum of x-ray wavelengths. The insertion device operating under these conditions is called a wiggler.

X-Ray Tubes. X-ray tubes are the most widely used source for the generation of x-rays. In these tubes, electrons are accelerated by a high electric potential (20–120 kilovolts). These electrons strike the target (anode) of the tube. The electrons decelerate as they pass through the electron clouds of the atoms. This phenomenon produces a continuous spectrum similar, but much less intense, to that of the synchrotron. In addition, some high energy electrons knock electrons out of the atomic orbitals of the atoms of the target material. When these orbitals are refilled by electrons, x-ray photons are generated. The resulting x-ray spectrum of intensity vs wavelength has a series of peaks known as characteristic lines. The wavelengths and intensities of these lines are dependent on the elemental composition of the target material (synchrotron radiation has no characteristic lines). The most intense peaks are the two lines known as $K_{\alpha 1}$ and $K_{\alpha 2}$ (or together known as K_{α}). The next most intense lines are the K_{β} lines.

Unlike for synchrotron radiation, the maximum intensity of x-rays from an x-ray tube is limited by how fast heat can be removed from the target to prevent its melting. In a conventional sealed tube, the target is stationary, relatively small, and must be continually cooled with water. In a rotating anode tube, the target is larger and is continually rotated so that the heat can be distributed over a larger surface. With such a tube the amount of heat, and hence,

the intensity of x-rays can be greater than in a stationary target tube.

The materials that are used as targets in x-ray tubes depend on the application.

Properties of X-Rays

X-rays impinging on an object interact primarily with the electrons in the object, ie, there is very little interaction with the atomic nuclei that comprise most of the mass of the object.

An x-ray photon can interact with an object in the following ways:

- (1) The x-ray photon does not interact and is transmitted through the object in the same direction that it impinged on the object. The energy of the photon does not change.
- (2) The x-ray photon is completely absorbed. No x-ray photon leaves the object. All of the energy of the x-ray photon is transferred to the electrons within the object.
- (3) The x-ray photon is absorbed and another x-ray photon of longer wavelength (lower energy) is produced. In this process an electron (or electrons) gain the lost energy. The new x-ray photon may travel in any direction from the site of the event. This phenomenon is known as incoherent, inelastic, or Compton scattering. Because there is a transfer of energy at a localized position within the object the incident x-ray photon can be considered to behave as a particle.
- (4) The x-ray photon produces an oscillating electric field in the object. This electric field in turn causes a photon of the same wavelength as the original photon to be generated. The resulting photon may leave the object at many different angles with respect to the incident photon. No energy is transferred to the object. This phenomenon is known as coherent, elastic, or Rayleigh scattering. The incident x-ray photon can be considered to behave as a wave and the phenomenon of diffraction takes place (see below).

All of these properties of x-rays are used to measure various properties of materials. X-ray applications can be placed into three categories based on which of the above phenomena are exploited. These categories are x-ray radiography, x-ray fluorescence spectrometry, and x-ray diffraction.

X-rays were discovered by Wilhelm Roentgen in 1895. Soon after their discovery Roentgen used x-rays to image the bones in a human hand. The x-rays transmitted through the hand (phenomenon 1) were recorded on photographic film. The bones or other dense objects absorb x-rays more than do soft tissues or less dense materials (phenomenon 2). This method of imaging the internal structure of an object is known as x-ray radiography.

X-ray fluorescence spectrometry consists of the measurement of the incoherent scattering of x-rays (phenomenon 3 above). It is used primarily to determine the elemental composition of a sample.

X-ray diffraction consists of the measurement of the coherent scattering of x-rays (phenomenon 4 above). X-ray diffraction is used to determine the identity of crystalline phases in a multiphase powder sample and the atomic and molecular structures of single crystals. It can also be used to determine structural details of polymers, fibers, thin films, and amorphous solids and to study stress, texture, and particle size.

Each of these categories of applications and the associated x-ray instruments are discussed in this article.

X-Ray Diffraction Principles

Interference of Waves. The coherent scattering property of x-rays is used in x-ray diffraction applications. Two waves traveling in the same direction with identical wavelengths, λ , and equal amplitudes (the intensity of a wave is equal to the square of its amplitude) can interfere with each other so that the resultant wave can have anywhere from zero amplitude to two times the amplitude of one of the initial waves. This principle is illustrated in Figure 1. The resultant amplitude is a function of the phase difference between the two initial waves.

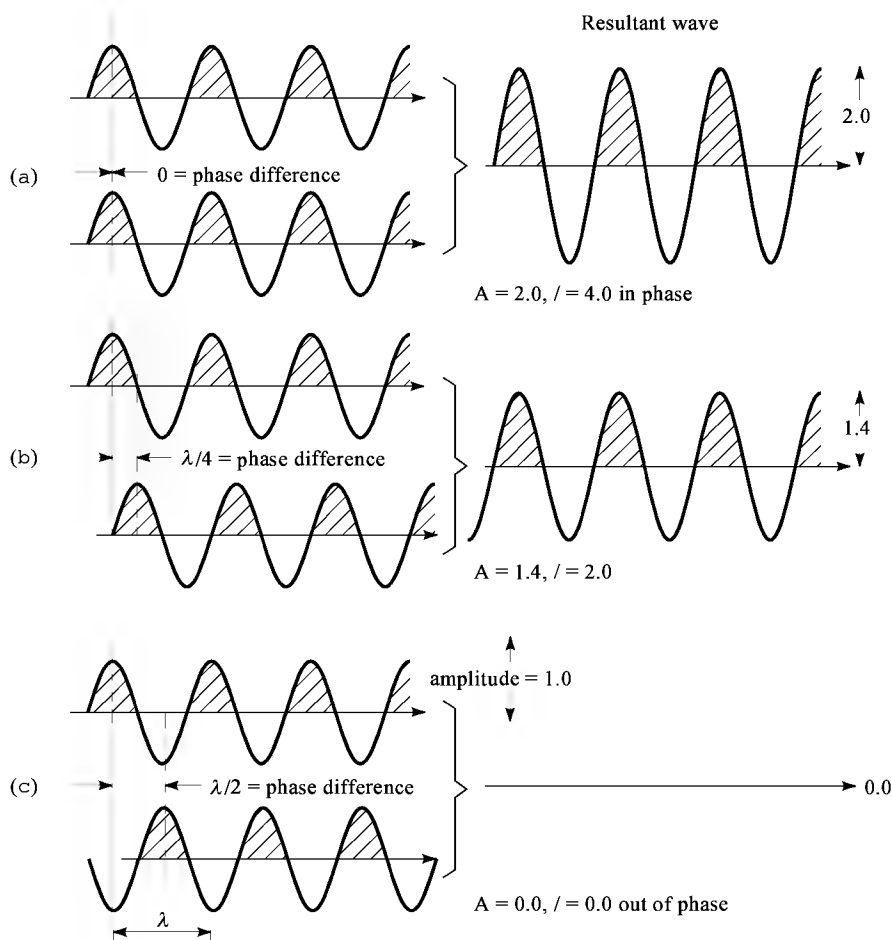


Fig. 1. Interference of two waves. (See text for explanation of a, b, c..)

In Figure 1a, the two waves have a zero phase difference and the resultant amplitude is twice that of each of the initial waves. In this case the waves are in phase with one another and the interference is "constructive."

In Figure 1b, the two waves have a phase difference of $1/4$ of the wavelength, $\lambda/4$, and the resultant amplitude is the square root of two times that

of each of the initial waves.

In Figure 1c, the two waves have a phase difference of one half of the wavelength, $\lambda/2$, and the resultant amplitude is zero. In this case the waves are out of phase with one another and the interference is "destructive."

In each of the three cases the resultant wave has the same wavelength, λ , as the initial waves.

Diffraction Patterns. A classical experiment in physics is the diffraction of light through slits or holes. If the size of the slits or holes are about the same size as the wavelength of the incident light, a diffraction pattern results. The diffraction pattern through a single slit is shown in Figure 2. The resulting pattern is a bell-shaped curve which is wider than the slit. The analog of the slit or hole in an x-ray diffraction experiment is an electron which is the entity that produces the scatter of an x-ray photon. An electron in an atom has a size on the same order of magnitude as the wavelength of the x-rays used in a diffraction experiment.

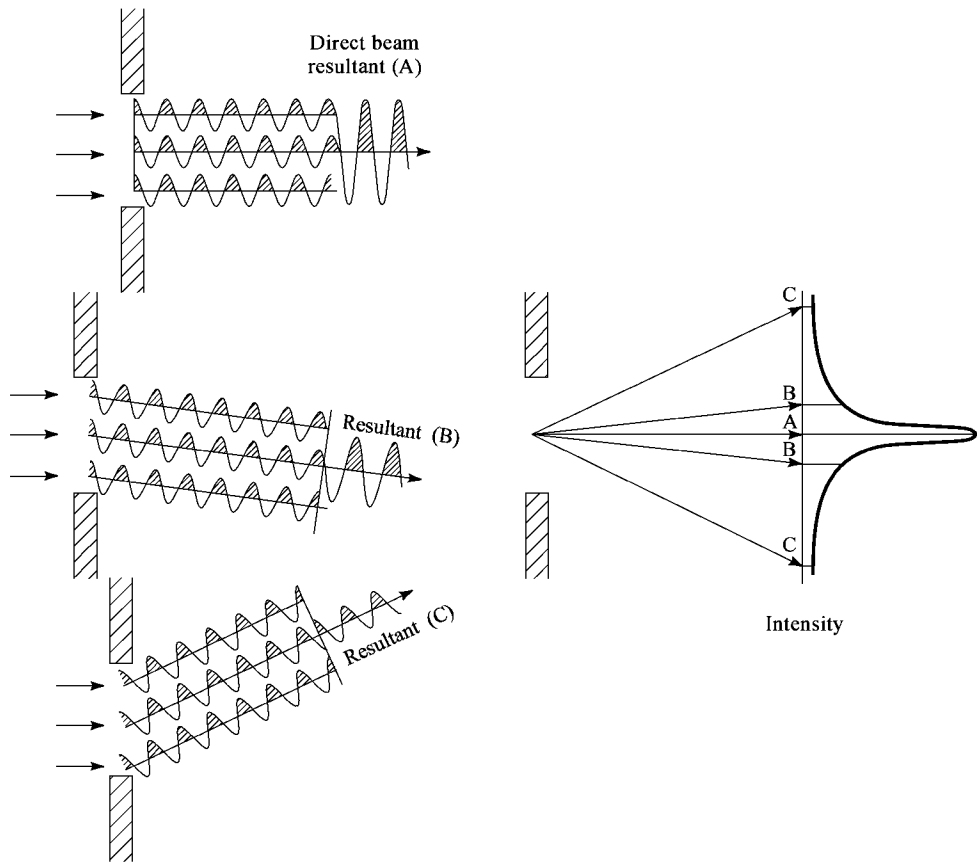


Fig. 2. Diffraction through a slit. Superposition of events **a**, **b**, **c**, from the same slit, plus all other possible angles of diffraction gives the resultant intensity profile.

Diffraction of x-rays from many atoms with many electrons results in a two-dimensional continuous pattern of peaks and valleys. Diffraction from a liquid or an amorphous (noncrystalline) solid results in a continuous pattern with few features, mainly broad peaks and valleys.

Unit Cells and Crystal Lattice. A single crystal, unlike a liquid or amorphous solid, is made up of identical units each of which has the same three-dimensional arrangement of atoms (and, hence, electrons). This basic unit is called a unit cell. It is a parallelepiped (a six-sided box whose opposite faces are parallel). The unit cell parameters are the lengths of the edges, a , b , and c , and the angles, α , β , and γ (Fig. 3). The edge lengths are usually expressed in angstroms ($1.0 \text{ \AA} = 10^{-10} \text{ m}$) and the angles are expressed in degrees. A crystal can be thought of as consisting of many of these unit cells which are stacked like bricks. Figure 4 shows a two-dimensional projection of a three-dimensional crystal structure. The positions of the atoms are shown as dots. The parallelograms are two-dimensional projections of the unit cells. The corners of the unit cells are the crystal lattice points. The crystal lattice consists of all of the lattice points in the crystal. Note that each unit cell has exactly the same configuration of atoms as every other unit cell.

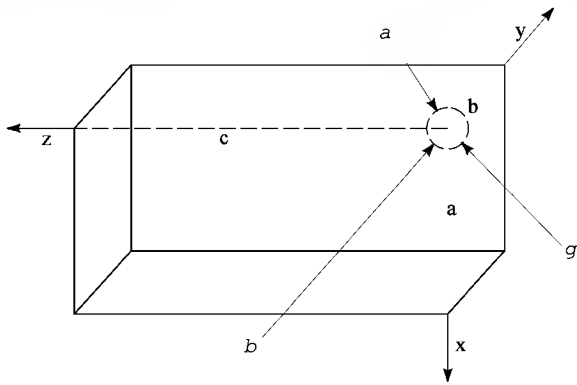


Fig. 3. Conventions for naming axes and angles for a unit cell.

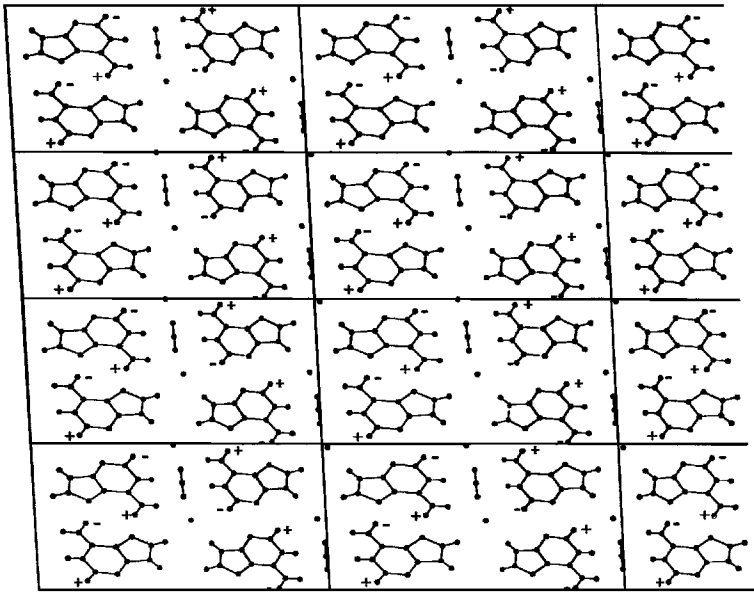


Fig. 4. Translational symmetry. For each atom in a unit cell there are corresponding atoms in neighboring unit cells.

Diffraction from a Crystal Lattice. An x-ray diffraction pattern from a single crystal has sharp peaks and regions of zero intensity between these peaks. An analogy is a diffraction pattern obtained by passing visible light through a series of equally spaced holes of uniform size (Fig. 5). In this case, each hole is analogous to a unit cell in the crystal.

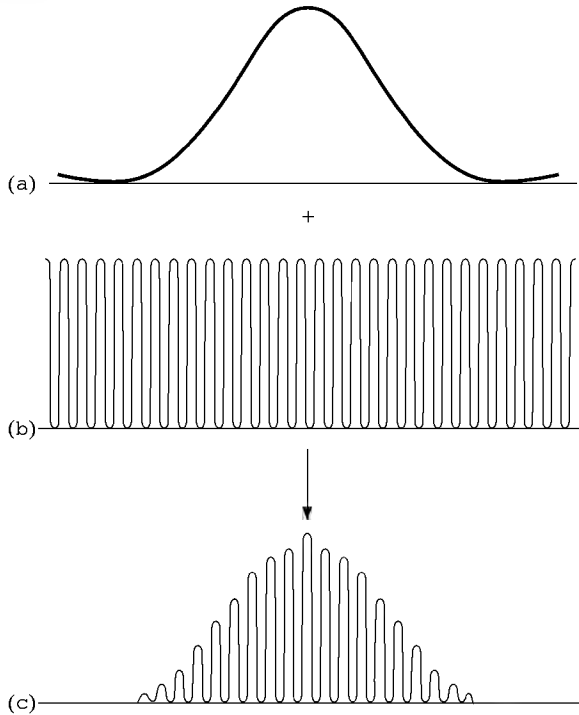


Fig. 5. Diffraction profile from (a) a single slit and (c) many slits. (b) The sampling region from many slits.

For a two-dimensional array of equally spaced holes the diffraction pattern is a two-dimensional array of spots. The intensity between the spots is very small. The crystal is a three-dimensional lattice of unit cells. The third dimension of the x-ray diffraction pattern is obtained by rotating the crystal about some direction different from the incident beam. For each small angle of rotation, a two-dimensional diffraction pattern is obtained.

In principle, it is possible to calculate the detailed three-dimensional electron density distribution in a unit cell from the three-dimensional x-ray diffraction pattern.

Bragg's Law. In 1913 W. L. Bragg showed that the positions of the discrete x-ray spots in the diffraction pattern can be explained by assuming that the diffracted x-ray photons behave as if they were "reflected" from certain families of equally spaced parallel planes passing through the crystal lattice. Figure 6 illustrates the principle known as Bragg's Law. The two paths, *A* and *B*, for the incident and diffracted beams differ in length by $2d \sin \Theta$ where *d* is the perpendicular distance between two adjacent parallel planes and Θ is equal to the angle between the incident beam and each plane and is equal to the angle between the diffracted beam and each plane. For constructive interference to take place this difference must be equal to a whole number of wavelengths. Bragg showed that diffraction takes place only if all the lattice points of the crystal are on the parallel planes.

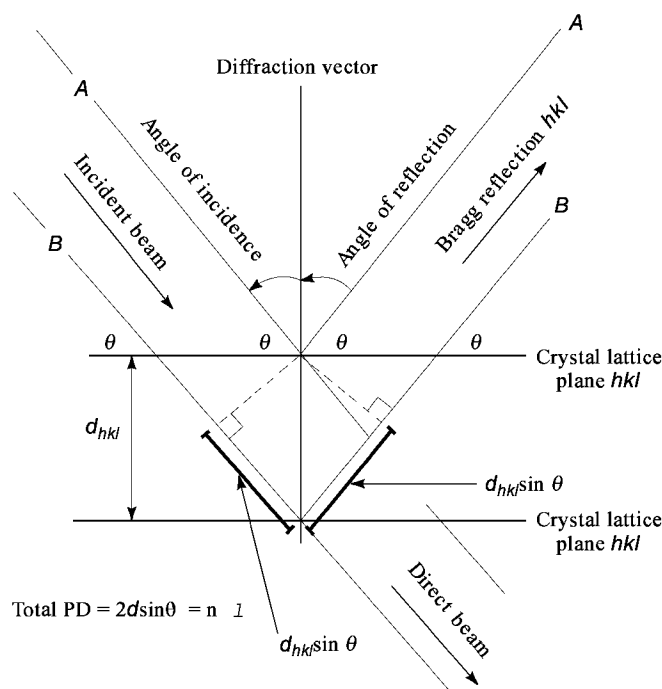


Fig. 6. Geometry of diffraction and its relationship to Bragg's Law.

Bragg's law is

$$n\lambda = 2d\sin\Theta$$

where n is an integer. An example of parallel planes which pass through all of the lattice points is shown in Figure 7.

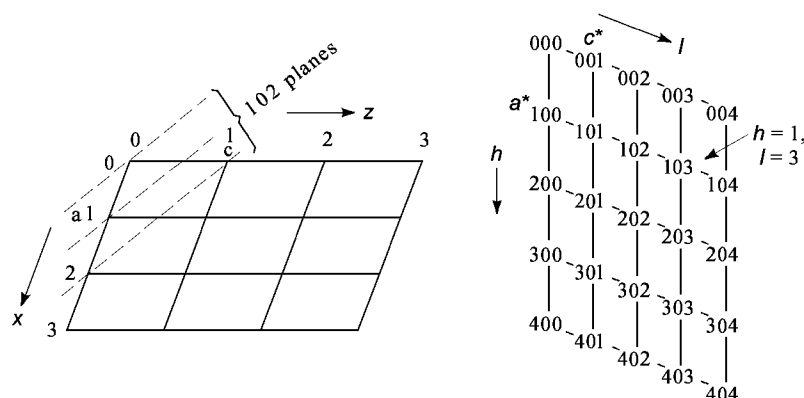


Fig. 7. Indices of reflections and the reciprocal lattice.

Because of Bragg's explanation of diffraction of x-rays from a crystal as being like reflections from families of planes, the diffraction spots are usually called "reflections." Each reflection is identified with three integer indices, h , k , and l . For the set of planes shown in Figure 7, the indices of the corresponding reflection are $h = 1$, $k = 0$, and $l = 2$.

Instruments for X-Ray Single-Crystal Diffraction

An x-ray single-crystal diffraction experiment consists of mounting a single crystal on a diffractometer, finding a few reflections, assigning indices to these reflections, determining the unit cell parameters and the orientation of the crystal on the diffractometer and then systematically measuring the intensities of all of the reflections. An x-ray single-crystal diffractometer consists of:

- (1) *An x-ray source.* Either synchrotron radiation or x-radiation from an x-ray tube is used (see above). The target materials for x-ray tubes for single-crystal diffraction experiments are usually copper or molybdenum. The resulting wavelengths are 15.418 nm and 7.1073 nm for CuK_α and MoK_α respectively.
- (2) *An x-ray monochromator.* A monochromator is a large single crystal (usually graphite) that is oriented so that a very intense reflection is directed toward the sample. All wavelengths are absorbed by the monochromator except a small range of wavelengths used for the diffraction experiment. Usually only the K_α characteristic radiation is used if an x-ray tube is the x-ray source.

The x-ray photons generated by the x-ray tube diverge from the target of the tube in all directions. Normally only those photons which are directed toward the monochromator and are diffracted from the monochromator toward the sample can be diffracted from the crystal and measured by the detector. Recently, graded multilayer devices have been used instead of single-crystal monochromators. The graded multilayers consist of 20 to 100 layers of alternating light elements (eg, silicon) and heavy elements (eg, tungsten). The layer thickness is graded from one side of the device to the other. Useful thicknesses are 1.5 to 10 nm. The devices are also curved so that they form a piece of a parabola (Fig. 8). A divergent beam of x-rays entering this device will exit as a parallel beam of monochromatic x-rays. The advantage compared to conventional monochromators is that the resulting intensity is several times larger and the parallel nature of the beam makes it possible to measure the intensities of reflections from crystals with large unit cell dimensions. A divergent beam causes reflections to overlap for samples with large unit cells (proteins).

- (1) *A goniometer.* The crystal is mounted in the center of a goniometer, which with computer control, can orient the single crystal in many different directions in three-dimensional space (Fig. 9).

(2) *An x-ray detector.* Two different types of detectors are commonly used.

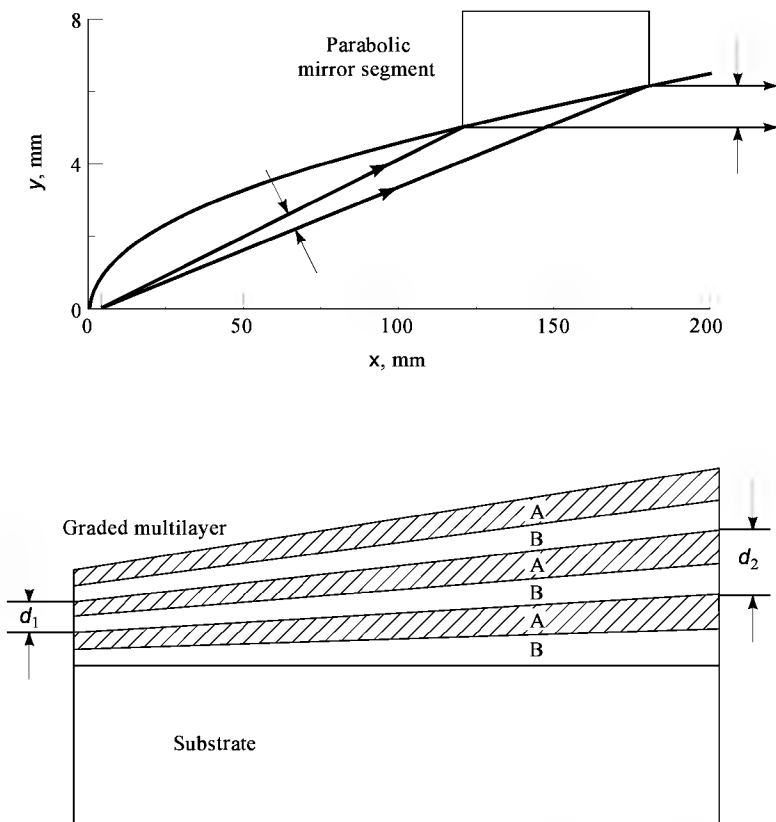


Fig. 8. A parabolically curved graded multilayer device. $\Delta y = 1 \text{ nm}$. $\Delta\alpha = 0.4^\circ$.

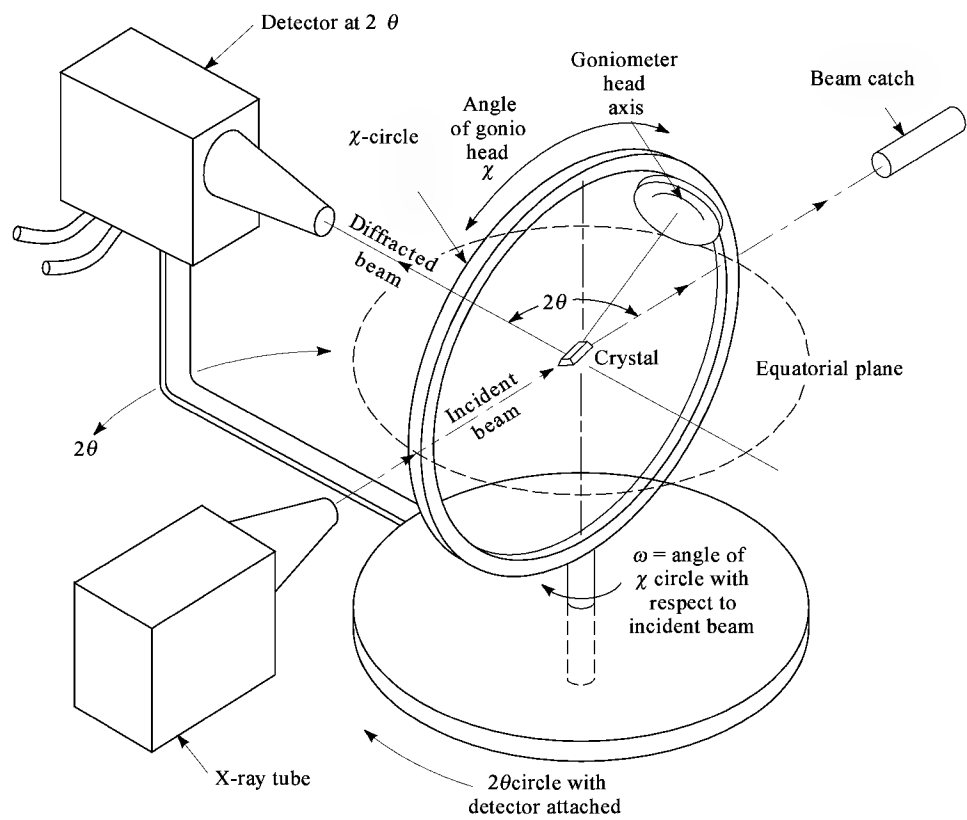


Fig. 9. A 4-circle single-crystal diffractometer system.

A single-reflection detector measures the intensity of one reflection at a time. It is necessary to orient the crystal and the detector (with the goniometer) with respect to the monochromatic incident x-ray beam for each reflection so that the Bragg condition is satisfied. A common single-reflection detector consists of a NaI crystal doped with Tl. The crystal is coupled to a photomultiplier tube that is followed by a voltage discriminator. This type of detector is called a scintillation detector. An x-ray photon hitting the NaI crystal is converted to many visible light photons, that are then multiplied by the photomultiplier tube, producing a voltage signal. For each signal that is in the correct range a count of one is added to the accumulated counts in the computer.

An x-ray area detector can be used to collect the intensities of many reflections at a time. The crystal must be oriented in many different settings with respect to the incident beam but the detector needs to be positioned at only a few positions to collect all of the data. A charge coupled device (CCD) is used as the area detector on the Siemens SMART single crystal diffractometer system. The SMART detector consists of a flat 6-cm circular phosphorescent screen that converts x-ray photons to visible light photons. The screen is coupled to a tapered fiber optics bundle which is then coupled to a one inch by one inch square CCD chip. The CCD chip has 1024×1024 pixels each of which stores an electrical charge proportional to the number of

x-ray photons detected at that point. When an x-ray image has been collected, it can then be converted from analog to digital and read into computer memory (Fig. 10).

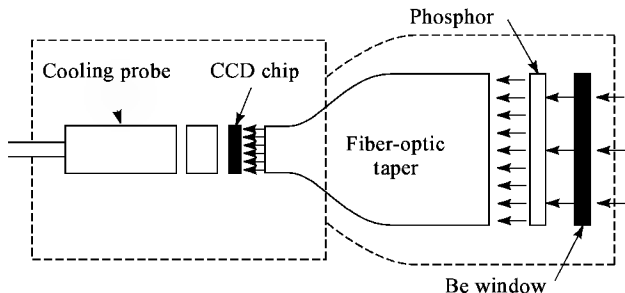


Fig. 10. Siemens SMART CCD Detector.

Typically it takes two to three days to collect a complete data set using a single-reflection detector. The new SMART diffractometer with its CCD detector can collect two or three data sets per day.

Figures 11 and 12 are diagrams of commonly used diffractometers.

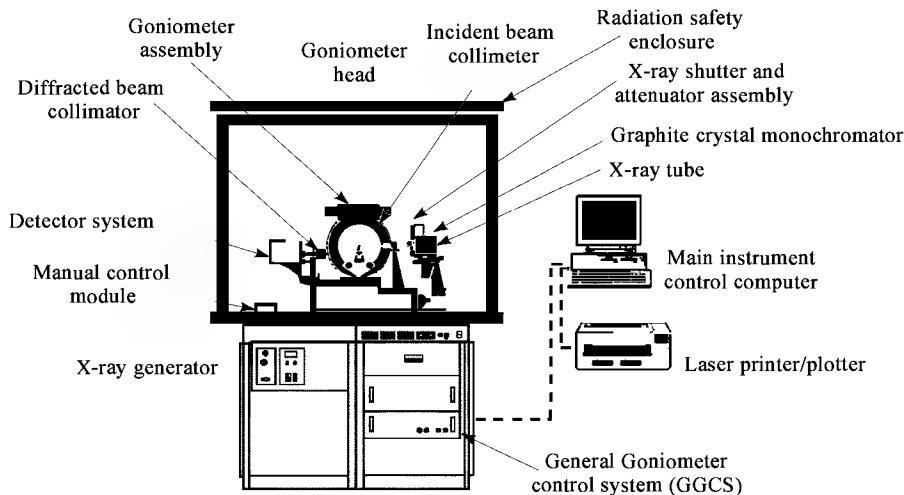


Fig. 11. Siemens P4 Single Crystal X-ray Diffractometer System.

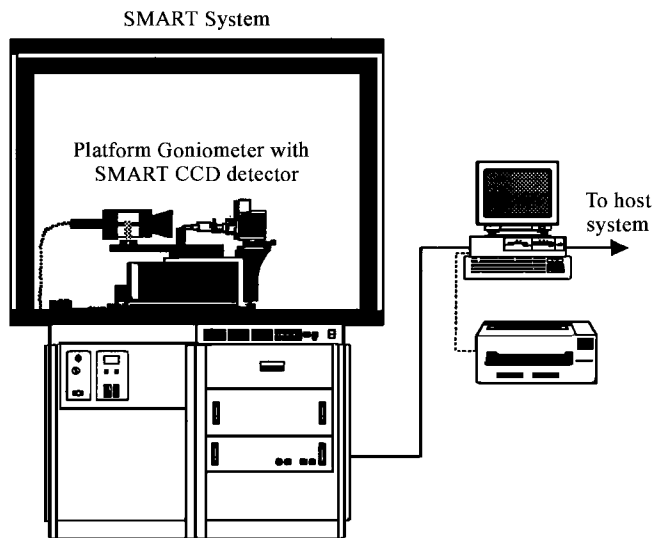


Fig. 12. Siemens SMART Single Crystal X-ray Diffractometer System.

Small-Molecule Single-Crystal Structure Determination

For the purposes of the following discussion small molecules are considered to have fewer than 200 nonhydrogen atoms. Macromolecules, primarily proteins, polynucleic acids, and viruses, can have many thousands of atoms.

A typical small-molecule structure determination using a SMART system proceeds in the following way:

Step 1. A single crystal with the largest dimension equal to about 0.01 mm to 0.3 mm is mounted on a glass fiber which in turn is mounted on a copper pin. The copper pin is placed on a goniometer head which in turn is placed on the goniometer (Fig. 13). The crystal is positioned manually in the center of the goniometer. In this position, the crystal is always in the center of the monochromated incident x-ray beam (whose diameter is about 1.0 mm).

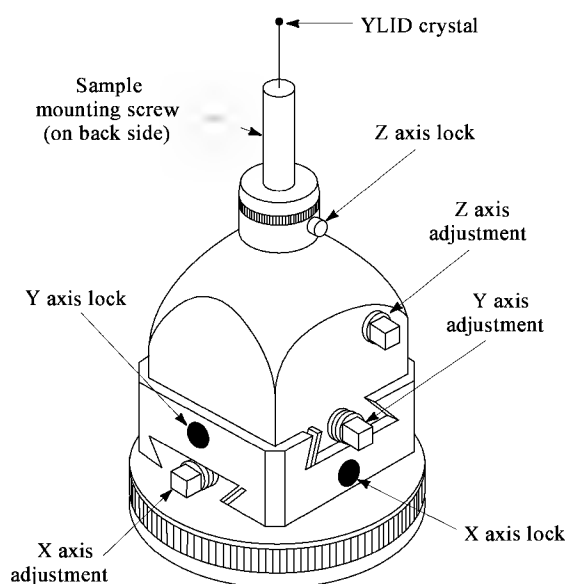


Fig. 13. Goniometer head with a mounted crystal.

Step 2. The computer opens a shutter, bathing the crystal in a monochromatic x-ray beam. The computer rotates the crystal for about one minute and the rotation diffraction image is stored on the detector and then read into the computer memory. When the operator examines the image and is confident that the sample is indeed a single crystal, the experiment can proceed.

Step 3. The computer collects about 45 frames of data. The crystal is rotated about the vertical axis for 0.3 degree for each frame. Usually the crystal is exposed to x-rays for about 5 seconds for each frame. The computer finds the centers of many reflections (typically 25 to several hundred) and determines indices for these reflections. It then determines the unit cell parameters and the orientation of the unit cells with respect to the diffractometer.

Step 4. The user may then search for a match of the measured unit cell parameters against all unit cell parameters that have been published. Currently, there are about 197,500 entries in the National Institute of Standards and Technology (NIST) Crystal Data File. An exhaustive search takes about one minute. Unit cell parameters are very definitive. Usually only one or a few hits are found and the appropriate literature reference(s) are listed. If no hits are found, the structure has not been previously reported.

The time for these four steps is typically about 15–20 min. It takes 1–2 h with the older single-reflection detector instruments. Because all reflections with weak, as well as moderate to strong intensities, are measured with the SMART system, the probability of obtaining incorrect unit cell parameters is less than for conventional instruments.

For small crystals, the 5 s per frame must be increased, but because so many reflections can be measured in a short period of time, it is possible to determine the unit cell parameters and identify crystals considerably smaller than was possible with conventional instruments. Currently (1996), the smallest sample for which this method has been used had dimensions of approximately 0.01 mm $\times$ 0.01 mm $\times$ 0.01 mm (about 2 n). As a result the SMART system is one of the most powerful tools for determining the identity of very small samples.

Step 5. If the crystal structure has not been previously reported, a complete data set is then collected automatically. The time for this operation is from 3 h for strongly diffracting crystals to 24 h for the weakly diffracting or small crystals.

At this point the crystal may be removed from the diffractometer making the instrument available to analyze another sample.

Step 6. The intensities are measured by the computer for each reflection in the dataset.

The techniques for determining and refining the atomic parameters obtained from a crystal structure are not described in detail in this article. For more details see the book by Glusker and Trueblood. The steps that the crystallographer takes to determine and refine the structure are outlined below.

Step 7. The computer determines the symmetry among the intensities of the reflections collected. The computer also looks for the absence or presence of intensity for certain classes of reflections. The computer determines which of 230 space groups the crystal belongs to from this information.

Step 8. To be able to calculate the three-dimensional electron density distribution of the crystal structure, it is necessary to know the phase-angle change from the incident to the diffracted beam for each reflection, as well as the intensity of each reflection. Unfortunately, the phase-angle information cannot be measured. This missing information leads to the so-called phase problem of x-ray crystallography. However, it is known that the electron density must be everywhere positive and that the electron density is distributed primarily at the positions of the atoms. From this information and the intensity of the reflections, very powerful computer programs have been written to solve the "phase problem." As a result, almost every structure with less than 200 nonhydrogen atoms in the unit cell is solved with these programs.

Step 9. After obtaining phase angles for many of the reflections (phase determination programs do not determine the phase angles of all of the reflections), a three-dimensional electron density map is generated. Because there are many reflections whose phase angles are not determined, this map has some false detail. The computer generates a list of the strongest peaks on this map.

Step 10. The crystallographer, using his or her knowledge of chemical information, such as bond distances and bond angles, analyzes the peak information and assigns labels and atomic types (eg, C, N, O, Si, Fe, etc) to those peaks which make good chemical sense.

Step 11. At this point a computer program refines the atomic parameters of the atoms that were assigned labels. The atomic parameters consist of the three position parameters (x , y , and z) for each atom. Also one or six atomic displacement parameters that describe how the atom is "smeared" (due to thermal motion or disorder) are refined for each atom. The atomic parameters are varied so that the calculated reflection intensities are made to be as nearly equal as possible to the observed intensities. During this process, estimated phase angles are obtained for all of the reflections whose intensities were measured. A new three-dimensional electron density map is calculated using these calculated phase angles and the observed intensities. There is less false detail in this map than in the first map.

The crystallographer assigns more labels and atomic types to more peaks and repeats the refinement. After a few cycles of refinement and assignment of atoms, all of the atoms, with the possible exception of some hydrogen atoms, are included in the model.

Step 12. Finally, the crystallographer analyzes the results, tabulates some critical information (eg, bond distances, bond angles, torsion angles, planarity of groups of atoms, etc), and prepares a report on the structure determination. The time for structure solution and refinement as outlined here is normally 2 to 8 h depending on the complexity of the structure.

Macromolecule Single-Crystal Structure Determination

The procedures for solving and refining macromolecular structures are similar to those for small molecules. The class of macromolecular structures solved consists mostly of proteins, some polynucleic acids, and some viral structures. These structures usually contain thousands of atoms.

These structures do not diffract as well as small molecules, and as a result, there are many weak reflections and data collection takes much longer than for small molecules. Also, the solution of the phase problem is more difficult and usually requires the collection of data sets with monochromatic radiation at several different wavelengths. Because of the much longer data collection times, area detectors are almost always used. Also because of the long

data collection times crystallographers like to use the very intense synchrotron sources where it is possible to collect all of the data in a few days instead of the weeks that may be required with x-ray tubes.

These structures tend to have a lot of disorder. It is rare that any hydrogens can be observed in the final electron density maps. In fact, many groups of atoms (water molecules and some side chains) may be so disordered that it is usually difficult to determine positions for these groups.

The Wealth of Information from Single-Crystal Determinations. The amount of information that is determined from a crystal structure experiment is much greater and more precise than for any other analytical tool for structural chemistry or structural molecular biology. Indeed, almost all of the structural information that has been determined for these two fields has been derived from x-ray single crystal diffraction experiments.

There are several databases that contain the structural information that has been published. The Cambridge Structural Database (CSD) contains information from studies of organic and organo-metallic compounds and metal-organic complexes. This database has almost 100,000 structures and is growing at the rate of 7000 entries per year. The Inorganic Crystal Structure Database (ICSD) has more than 3000 entries and is increasing at about 1200 entries per year. The Protein Databank (PDB) has structural information for about 1800 macromolecular structures (1993). The National Research Council of Canada maintains a Metals Crystallographic Data File (CRYSTMET) which contained about 6000 entries in 1987.

There are very sophisticated programs to search these databases. The databases are used by chemists to design new chemical compounds and by drug companies to design new drugs for a great variety of diseases. Of special interest is the design of drugs that fit into the active sites of proteins.

Instruments for Powder Diffraction

Bragg-Brentano Powder Diffractometer. A powder diffraction experiment differs in several ways from a single-crystal diffraction experiment. The sample, instead of being a single crystal, usually consists of many small single crystals that have many different orientations. It may consist of one or more crystalline phases (components). The size of the crystallites is usually about 1–50 μm in diameter. The sample is usually prepared to have a flat surface. If possible, the experimenter tries to produce a sample that has a random distribution of crystallite orientations.

See Figure 14(a) for the design of the most commonly used powder diffractometers. The instruments are designed so that a divergent beam of x-rays impinges on the sample. A convergent beam of x-rays is diffracted from the sample and passes through a narrow slit. The normal to the flat surface of the sample bisects the angle between the incident and diffracted beams. Usually a single-crystal monochromator is placed after the diffraction slit to remove x-ray photons of unwanted wavelengths. A conventional scintillation detector (see above) is usually used to measure the intensity of the diffracted beam. This kind of powder diffractometer is called a Bragg-Brentano diffractometer. This type of diffractometer gives best results if the sample is very flat.

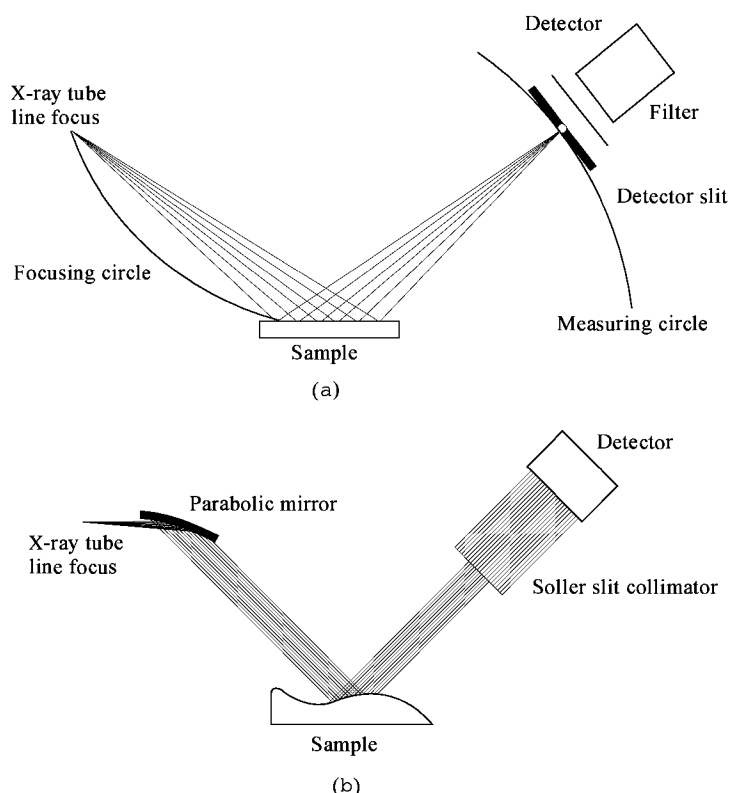


Fig. 14. Focusing schemes in powder diffraction: (a) conventional para-focusing Bragg-Brentano diffractometer; (b) parallel-beam diffractometer using a parabolic graded multilayer device.

Graded Multilayer Device. Recently, a new diffractometer design takes advantage of the graded multilayer device. Figure 8 shows the principle of this device. With this device, parallel incident and diffracted beams are produced. No diffracted beam monochromator is necessary. The advantage over the Bragg-Brentano design is that the sample does not have to be flat (Fig. 14(b)).

Applications. The most common application for the Bragg-Brentano and graded multilayer powder diffractometers is to measure the 2θ -values (2θ is the angle between the incident and diffracted x-ray beams; see Bragg's law above) and intensities for the peaks in a powder diffraction pattern. A typical experiment consists of mounting the specimen in the center of the instrument and then slowly rotating the specimen and detector (the detector moves at twice the speed of the specimen) so that the diffraction angle (2θ) increases. While the detector and specimen rotate, the x-ray photons that reach the detector are counted by the computer. Figure 15 is an example of a typical powder pattern. This pattern is analyzed by the computer in various ways.

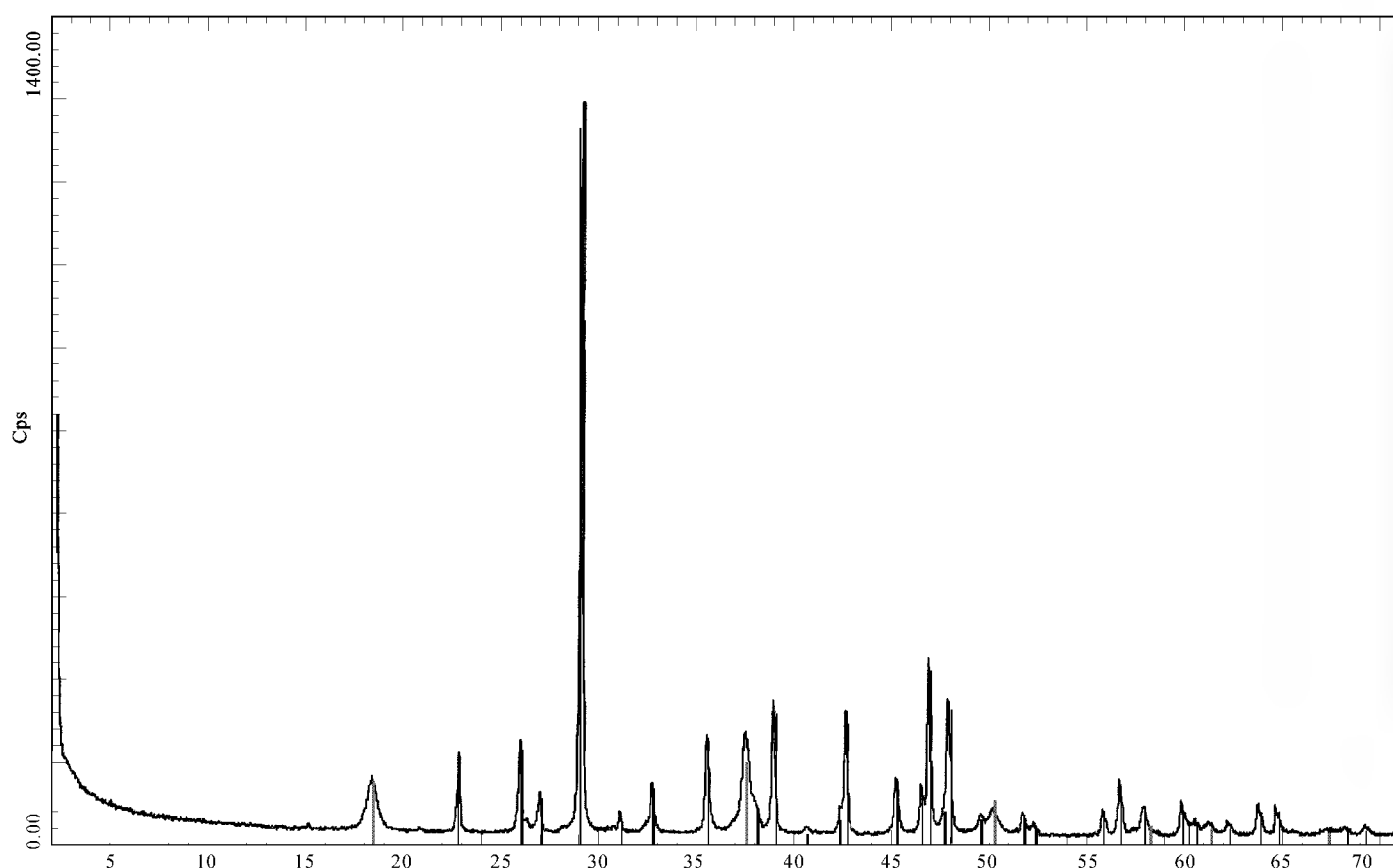


Fig. 15. A typical powder pattern with three phases: Calcite (—), Aragonite (—) and Brucite (~~~~). The lines below the peaks are the powder lines from the ICDD Powder Diffraction File.

Search—Match. The computer identifies which crystalline phases (components) match the unknown pattern by using a file of known powder patterns maintained by the International Center for Diffraction Data (ICDD). The Powder Diffraction File contains interplanar d -spacings [$d = \lambda / (2\sin\Theta)$] and intensities of the diffraction maxima for each crystalline powder pattern submitted to the ICDD. Currently there are about 65,000 patterns in the file. Current search-match programs can successfully identify up to seven components in an unknown pattern. A typical diffraction pattern of an unknown sample and the components identified by the computer search-match program is shown in Figure 15.

Quantitative Phase Analysis. Once the identity of the components in a sample are known, it is possible to determine the quantitative composition of the sample. There are several different methods for doing a quantitative analysis, but the most reliable method is to use mixtures of known composition as standards. The computer can determine quantitatively the relative amounts of each component in the unknown sample. For accurate calculations of relative amounts in the unknown sample, it is necessary that the sample and standards have uniform distributions of crystallites. Often the sample and standards are rotated during data collection to provide a more even distribution of crystallites which diffract.

Indexing and Lattice Parameter Determination. From a powder pattern of a single component it is possible to determine the indices of many reflections. From this information and the 2Θ -values for the reflections, it is possible to determine the unit cell parameters. As with single crystals this information can then be used to identify the material by searching the NIST Crystal Data File (see "Small Molecule Single Structure Determination" above).

Structure Determination from a Powder Pattern. In many cases it is possible to determine atomic positions and atomic displacement parameters from a powder pattern. The method is called the Rietveld method. Single-crystal structure determination gives better results, but in many situations where it is impossible to obtain a suitable single crystal, the Rietveld method can produce adequate atomic and molecular structures from a powder pattern.

Crystallite Size. From the width of the peaks the computer can determine the size of the crystallites in the sample. The smaller the crystallite size, the broader are the diffraction peaks. This kind of analysis is important for determining particulate size of certain materials (eg, silica) where a range of crystallite size may be a health hazard if inhaled into the lungs.

Measurement of Residual Stress and Strain. The displacement of the 2Θ -value of a particular line in a diffraction pattern from its nominal, nonstressed position gives a measure of the amount of stress retained in the crystallites during the crystallization process. Thus metals prepared in certain ways (eg, cold rolling) have stress in their polycrystalline form. Strain is a function of peak width, but the peak shape is different than that due to crystallite size. Usually the two properties, crystallite size and strain, are determined together by a computer program.

Percent Crystallinity. For samples that consist of a mixture of crystalline and amorphous material, it is possible to determine the percent of crystallinity by measuring the integrated intensity of sharp Bragg reflections and the integrated intensity of the very broad regions due to the amorphous scattering.

Texture Analysis. For many rigid polycrystalline materials (eg, metals) an analysis of the orientation of the crystallites is important to understanding the mechanical properties of the sample. In a texture analysis, the sample must be oriented in many different directions with respect to the incident beam. Special texture attachments which provide the necessary tilting of the sample are required. In a pole figure analysis, the intensity of one specific diffraction line (eg, the 200 reflection) is measured many times at a given 2Θ -value (the detector is held fixed). The measurements are made at different settings of the sample by rotating about two different axes (called ϕ and χ). From pole figures for several different reflections the computer can generate an orientation distribution function (ODF) which contains all of the information concerning the three dimensional distribution of the orientation of the crystallites.

Instruments for Special Applications

Because so many new and exotic materials are being manufactured, special x-ray instruments have been designed to measure properties of these materials.

X-Ray Reflectometer. Thin films can be deposited on single-crystal substrates. Such materials have special properties (eg, YBCO deposited on silicon is used as a superconductor which can detect very small magnetic fields). To be able to measure the thickness of thin layers, an x-ray reflectometer has been designed. Although this instrument measures the reflected x-ray photons rather than diffracted photons, the design is very similar to that of a conventional powder diffractometer. As with visible photons, total reflection can take place from a surface if the angle between the incident and reflected beams is small enough. Because the wavelengths of x-rays are very small, total reflection occurs at angles on the order of several tenths of a degree. A diagram of an x-ray reflectometer is shown in Figure 16. With this instrument it is possible to measure the thickness of several layers deposited on top of one another. It is also possible to measure the roughness of each layer. Typical measured thicknesses are tens of nanometers and typical

measured roughnesses are several tenths of nanometers.

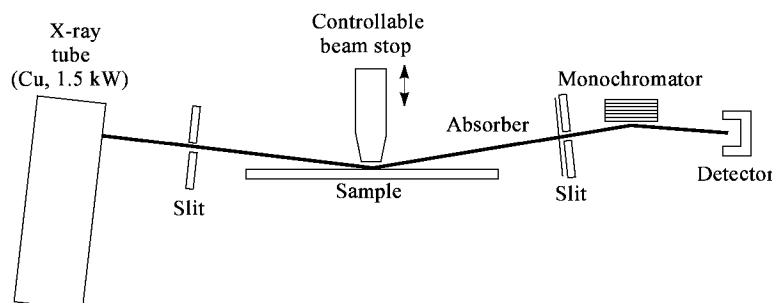


Fig. 16. Diagram of the Siemens X-ray reflectometer.

Position Sensitive Detectors. By replacing the scintillation detector in a conventional powder diffractometer with a Position Sensitive Detector (PSD), it is possible to speed data collection. For each x-ray photon received a PSD records the angle at which it was detected. Typically, a conventional scintillation detector records x-ray photons in a range of a few hundredths of a degree at a time. A PSD can measure many degrees (in 2θ) of a powder pattern simultaneously. Thus, for small samples, data collection, which could require hours with a conventional detector, could take minutes or even seconds with a PSD.

Area Detectors. A two-dimensional or area detector attached to a powder diffractometer can greatly decrease data collection time. Many diffraction applications require so much time with a conventional detector that they are only feasible if an area detector is attached to the instrument. The Siemens General Area Detector Diffraction System (GADDS) uses a multiwire area detector (Fig. 17). This detector measures an x - and a y -position for each x-ray photon detected. The applications follow.

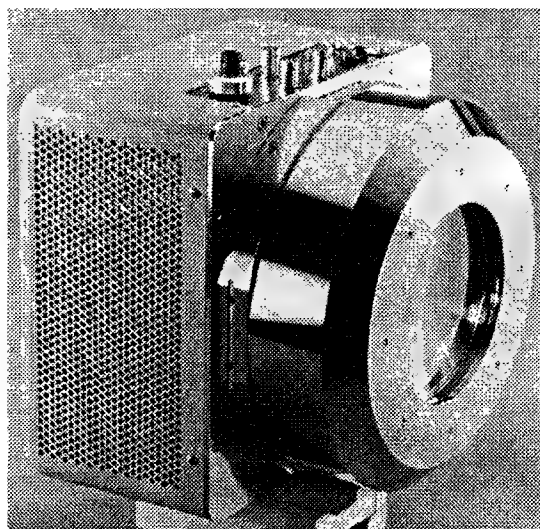


Fig. 17. Siemens HI-STAR multiwire proportional detector.

Texture Analysis with GADDS. With a conventional detector, a data collection for a pole figure analysis with a powder diffractometer with a texture attachment could take 12 h or more. With an area detector, it is possible to collect enough data for several pole figures (required for an ODF analysis) in a few minutes.

Small Angle Scattering (SAS). Diffraction at small angles gives information about large scale structure (d -spacings up to 90 nm). Usually the intensity at such small angles is weak and the ability of the area detector to collect complete diffraction rings makes this application feasible. SAS is used to characterize polymeric or fiber materials.

Polymer or Fiber Diffraction. Polymers and fibers are often ordered in one or two dimensions but not ordered in the second or third dimension. The resulting diffraction patterns have broad diffuse diffraction maxima. The ability to collect two dimensional images makes it possible to collect and analyze polymer and fiber diffraction patterns.

Microdiffraction. By concentrating the incident x-ray beam on a small portion of a sample it is possible to get a complete diffraction pattern of very small regions of a sample. Of course, the intensity from such small regions is weak and an area detector that can collect a large portion of the diffraction pattern at one time makes this application practical. A typical region size is about 50 μm in diameter.

X-Ray Fluorescence Spectrometry

X-ray fluorescence spectrometry is a technique for measuring the elemental composition of samples. The basis of the technique is the relationship between the wavelength or energy of the emitted incoherently scattered x-ray photons and the atomic number of the element. This relationship established in 1913 is

$$E = 1.24 \times 10^{-1} \lambda^{-1} = K(Z - s)^2$$

where E is the energy in kiloelectron-volts (keV), λ is the wavelength in nanometers, Z is the atomic number of an element and K and s are constants that depend on the spectral series of the emission line in question. When an atom is bombarded with x-ray photons of sufficient energy, an inner-orbital electron may be displaced leaving the atom in an excited state. As an example, if a molybdenum atom were bombarded in this way, a vacancy in the innermost shell (K shell) might be created by removal of one or the $1s$ orbital electrons (see Fig. 18). The energy to remove the K shell electron from the atom is, in the case of molybdenum, about 20 keV. The atom can return to the ground state by transference of a higher orbital electron into the K shell vacancy (the resulting higher level vacancy is filled by an electron from a still higher level and so on). In so doing, the difference in energy between the electron in the K shell and the energy of the electron in the shell from which it came is emitted as an x-ray photon. For molybdenum, an L shell electron transferring to the K shell causes an x-ray photon with an energy of 17.6 keV to be emitted. For the L to K shell transition the emitted x-ray photon is called a K_α photon. L series x-ray photons are produced when L shell vacancies are filled with higher level electrons. The fluorescent yield is approximately one for the high atomic number elements and may be as small as 0.01 for the low atomic elements (eg, Be, Na, Mg, and Al). Most commercially available x-ray spectrometers can measure x-ray photons with wavelengths in the range of about 0.02–8.0 nm (60–0.18 keV) which means that the K series can be

measured for atomic numbers 4(Be) to 71(Lu) and the *L* series from 25(Mn) to 92(U). Each element produces a fluorescence spectrum of intensity vs wavelength which is characteristic of that element.

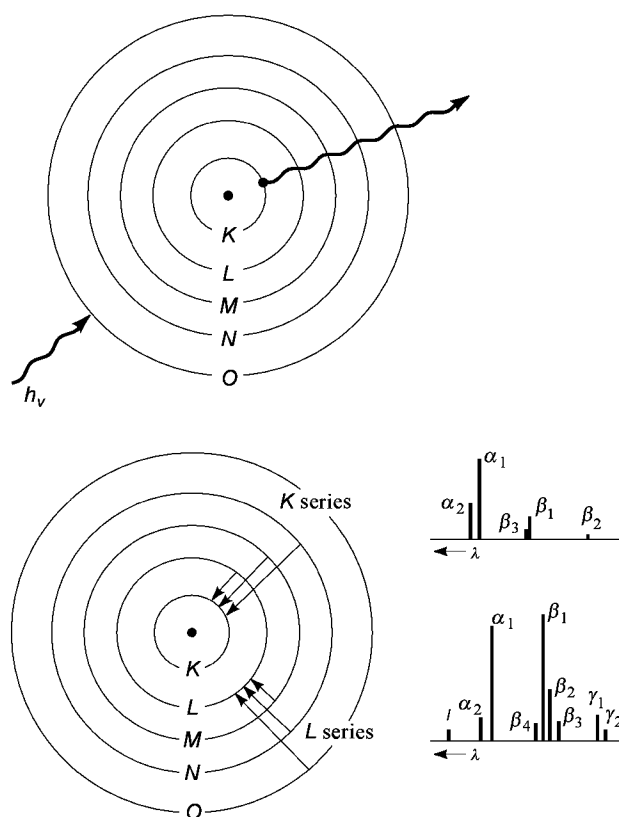


Fig. 18. Production of x-ray fluorescence *K* and *L* series from molybdenum atom.

X-Ray Spectrometers. An x-ray spectrometer is an instrument that measures the fluorescence spectra of samples. The associated computer software then determines the qualitative and quantitative elemental composition of the samples from the resulting spectra.

The primary x-ray source is usually a sealed tube with a Rh anode. The tube usually has an end window so that the sample is as close to the anode as possible. This arrangement ensures that the surface of the sample is irradiated with a beam with nearly uniform intensity.

Wavelength Dispersive Spectrometer. A wavelength dispersive spectrometer uses an analyzing crystal to measure the spectrum. The analyzing crystal is a single crystal positioned so that a specific Bragg reflection can be used to measure the intensity of the various wavelengths of the fluorescence spectrum. The analyzing crystal is continuously rotated. At each position of the crystal, x-ray photons with only a narrow band of wavelengths are diffracted in to a detector. The detector is usually a scintillation detector like that used in a conventional powder diffractometer. The detector is rotated at twice the speed of the analyzing crystal (see Fig. 17) so that the Bragg condition (Fig. 6) is maintained for each wavelength.

Because such a wide range of wavelengths must be measured (see above), no single analyzing crystal is best for all elements. The 1 0 0 reflection of LiF (*d* spacing = 0.403 nm) is useful for measuring the K_{α} lines for elements K to Xe. A pentaerythrite (PET) *d* spacing = 0.874 nm) analyzing crystal is useful for measuring the K_{α} lines for elements Al to Ti. Special multilayer materials (eg, alternating layers of W and C; *d* spacing = 16.0 nm) are useful for the K_{α} lines for the lowest atomic number elements. Other analyzing crystals are used for measuring other ranges of wavelengths.

Energy Dispersive Spectrometer. An energy dispersive spectrometer uses a detector that can measure the energy of each detected x-ray photon. The detector is usually a lithium-drifted silicon [Si(Li)] detector which is a proportional detector of high intrinsic resolution. The resolution of this type of detector is typically 160–180 eV, compared to 20–200 eV for the wavelength dispersive systems. A multichannel analyzer is used to sort the arriving pulses into memory locations corresponding to the energy of the x-ray photons. No analyzing crystal is needed for the energy dispersive systems. The complete spectrum is collected with no mechanical movement of the spectrometer.

Energy dispersive spectrometers can, in general, collect the spectrum faster and are less expensive than the more sophisticated wavelength dispersive spectrometers. However, they do not have the resolution and cannot separate closely spaced lines as easily as the wavelength dispersive spectrometers.

Both the wavelength dispersive and energy dispersive spectrometers are well suited for qualitative analysis of materials. Each element gives on the average only six emission lines. Because the characteristic x-ray spectra are so simple, the process of allocating atomic numbers to the emission lines is relatively simple and the chance of making a gross error is small.

The great flexibility and range of the two types of x-ray spectrometers make them ideal for quantitative analysis. The precision of a well-designed x-ray spectrometer is typically about 0.1%. Systematic errors in quantitative x-ray spectrometry arise mainly from absorption and specimen related phenomena, including particle size and heterogeneity. Although these effects are complex, computer programs have been developed for handling them. Relative amounts of the elements in a sample can usually be determined with an accuracy of a few tenths of a percent.

X-Ray Radiography

X-ray imaging tests are widely used to examine interior regions of metal castings, fusion welds, composite structures, and brazed components. Radiographic tests are made on pipeline welds, pressure vessels, nuclear fuel rods, and other critical materials and components that may contain three-dimensional voids, inclusions, gaps or cracks that are aligned so that the critical areas are parallel to the x-ray beam. Since penetrating radiation tests depend upon the absorption properties of materials on x-ray photons, the tests can reveal changes in thickness and density and the presence of inclusions in the material.

X-ray fluoroscopy is used for direct on-line examination. A fluorescent screen is used to convert x-ray photons into visible light photons. A television camera receives the visible image and displays it on a television screen (see Fig. 19). This type of system is used for security screening of carry-on luggage at airports.

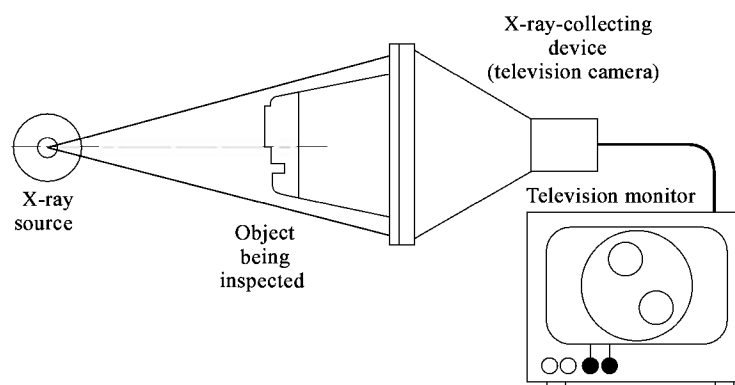


Fig. 19. Basic hardware for x-ray fluoroscopic inspection.

As in medical x-ray imaging, computerized tomography (CT) can reveal the details of the internal structure of complex objects. Many detectors are used to measure the transmittance of x-rays along many lines through the object. A computer uses this information to produce an image of the internal structure of a slice of the object. One of the methods under investigation by the Federal Aviation Administration is a computerized tomographic scanning system for automatically detecting bomb materials in bags loaded on to airplanes.

Applications Using Synchrotron Radiation

Because of the unique features of the x-ray radiation available at synchrotrons, many novel experiments are being conducted at these sources. Some of these unique features are the very high intensity and the brightness (number of photons per unit area per second), the nearly parallel incident beam, the ability to choose a narrow band of wavelengths from a broad spectrum, the pulsed nature of the radiation (the electrons or positrons travel in bunches), and the coherence of the beam (the x-ray photons in a pulse are in phase with one another). The applications are much more diverse than the applications described in this article. The reader may wish to read the articles in the Proceedings of the Materials Research Society listed in the bibliography.

Just two of the many experiments being conducted that are of special interest to chemists are described here.

Extended Absorption Fine Structure (EXAFS). As indicated above it is possible to obtain very detailed structural information from x-ray diffraction of single crystals. For samples which are not single crystals it is possible to obtain structural information around the site of certain heavy atoms. The procedure is called Extended Absorption Fine Structure (EXAFS). As indicated in the section, "X-Ray Fluorescence Spectrometry," when an x-ray photon of a specific wavelength impinges on an atom of a given type (eg, Fe), the incident photon may be absorbed by the atom and an x-ray photon with longer wavelength is emitted. The intensity of the transmitted x-rays at this specific wavelength decreases correspondingly. The resulting absorption spectrum of intensity vs wavelength will have a sharp peak at this specific wavelength. There is, however, a fine structure to this absorption spectrum which is dependent on the chemical structure of the atoms bonded to the atom of interest. By measuring the absorption fine structure spectrum of intensity vs wavelength, one can determine this bonding structure. Because of the large intensities and fine wavelength tunability of the incident x-ray photons available at synchrotrons, these measurements can be routinely performed, but they are difficult or impossible to perform using conventional x-ray sources. Much information about the structures of catalysts have been obtained using EXAFS.

Laue Method for Macromolecule X-Ray Diffraction. As indicated above it is possible to determine the structures of macromolecules from x-ray diffraction; however, it normally takes a relatively long period of data collection time (even at synchrotrons) to collect all of the data. A new technique, the Laue method, can be used to collect all of the data in a fraction of a second. Instead of using monochromated x-rays, a wide spectrum of incident x-rays is used. In this case, all of the reflections that are diffracted on to an area detector are recorded at just one setting of the detector and the crystal. By collecting many complete data sets over a short period of time, the Laue method can be used to follow the reaction of an enzyme with its substrate. This technique cannot be used with conventional x-ray sources.

BIBLIOGRAPHY

"X-Ray Fluorescence Spectrography" in *ECT* 1st ed., Vol. 15, pp. 176–185, by W. Parrish, Philips Laboratories; "X-Ray Analysis" in *ECT* 2nd ed., Vol. 22, pp. 438–467, W. Parrish, International Business Machines Corporation; "X-Ray Technology" in *ECT* 3rd ed., Vol. 24, pp. 678–708, by Ronald Jenkins, Philips Electronic Instruments, Inc.

General References

The following are general references for the topics discussed in this article.

X-ray Small Molecule Crystallography and Macromolecular Crystallography

J. P. Glusker and K. N. Trueblood, *Crystal Structure Analysis*, 2nd ed., Oxford University Press, New York, 1985.

J. P. Glusker, M. Lewis, and M. Rossi, *Crystal Structure Analysis for Chemists and Biologists*, VCH Publishers Inc., New York, 1994.

C. P. Brock, B. M. Craven, B. A. Frenz, E. Gabe, R. A. Sparks, C. E. Strouse, K. N. Trueblood, and B-C. Wang, *Structure Analysis by X-ray Crystallography*, 3rd rev., American Crystallographic Association, Buffalo, New York, 1995.

X-Ray Powder Diffraction

H. P. Klug and L. E. Alexander, *X-ray Diffraction Procedures for Polycrystalline and Amorphous Materials*, 2nd ed., John Wiley & Sons, Inc., New York, 1974.

B. D. Cullity, *Elements of X-ray Diffraction*, 2nd ed., Addison-Wesley Publishing Company, Inc., Menlo Park, Calif., 1978.

R. Jenkins and R. L. Snyder, *Introduction to X-ray Powder Diffractometry*, John Wiley & Sons, Inc., New York, 1996.

X-Ray Fluorescence Spectrometry

R. E. Van Grieken and A. A. Markowicz, *Handbook of X-ray Spectrometry, Methods and Techniques*, Marcel Dekker, Inc., New York, 1993.

Computerized Tomography

R. Gordon, R. Bender, and G. T. Herman, *J. Theor. Biol.* **29**, 471–481 (1970).

A. M. Cormack, *Phys. Med. Biol.* **18**, 195–207 (1973).

U.S. Pat. 3,778,614, G. N. Hounsfield, *Method and Apparatus for Measuring X- or γ -ray Absorption or Transmission at Plural Angles and Analyzing the Data* (Dec. 11, 1973).

Synchrotron Radiation Applications

H. Winick and J. Doniach, eds., *Synchrotron Radiation Research*, Plenum Press, New York, 1980.

D. L. Perry, N. D. Shinn, R. L. Stockbauer, K. L. D'Amico, and L. J. Terminello, eds., *Applications of Synchrotron Radiation Techniques to Materials Science*, Materials Research Society Symposium Proceedings, vol. 307, 1993.

L. J. Terminello, N. D. Shinn, G. E. Ice, K. L. D'Amico, and D. L. Perry, eds., *Applications of Synchrotron Radiation Techniques to Materials Science II*, Materials Research Society Symposium Proceedings, vol. 375, 1994.

L. J. Terminello, S. M. Mini, H. Ade, and D. L. Perry, eds., *Applications of Synchrotron Radiation Techniques to Materials Science III*, Materials Research Society

Symposium Proceedings, vol. 437, 1996.

Robert A. Sparks
Siemens Analytical X-Ray Systems, Inc.



YEASTS

Yeasts are probably the oldest cultivated plants: their use dates to 3000 BC or earlier. They are eukaryotic organisms, ie, they have a defined nucleus surrounded by a nuclear membrane and organized cytoplasmic organelles (mitochondria, peroxisomes, vacuoles, etc). The recognition of yeasts as a coherent group of microorganisms and of their role in fermentation generally preceded their taxonomic classification. They are fungi that exist as single cells during at least some part of their life cycle, they have no photosynthetic ability, and are not motile. All the major groups of fungi, *Basidiomycetes*, *Ascomycetes*, *Phycomycetes*, and the *Fungi imperfecti*, include species that have adopted a yeast phase. Yeasts usually reproduce vegetatively by budding or fission, but parasexual and sexual reproduction occur as well in most yeasts. The most extensive and authoritative texts on yeast taxonomy and identification are References 1 and 2. Among those listed, the genus *Saccharomyces* is of greatest practical and economic importance for the baking, beer, and wine industries, as well as for the production of biomass. This article emphasizes the biology and industrial use of *Saccharomyces cerevisiae*. Owing to its importance in the study of basic cell biology, reviews of *Saccharomyces* genetics and molecular biology appear regularly; References 3 and 4 contain the most extensive reviews of yeast biology. Other yeasts participate in alcoholic fermentations and occur as food-spoilage organisms or pathogens.

Morphology, Reproduction, and Life Cycle

A single yeast cell is usually spherical to ellipsoidal in shape but may be cylindrical, ogival, pyramidal, or apiculate. Often the shape reflects the specific site of bud formation. Apiculate or lemon-shaped forms arise from bipolar budding, ie, buds form at the ends of the long polar axis of an elliptical cell. Yeast cell size varies. For ellipsoidal cells, eg, of *S. cerevisiae*, the length of the long axis may be 5–10 μm and of the short axis 3–5 μm (Fig. 1). The volume of such a cell may be 40 μ^3 with a mass of 10 pg and a density of 1.01–1.03 g cm^3 (5). Yeasts may form a pseudomycelium consisting of single or branched chains of cells if daughter cells do not separate from their mothers; starvation for nitrogen induces such a morphology in *S. cerevisiae*.

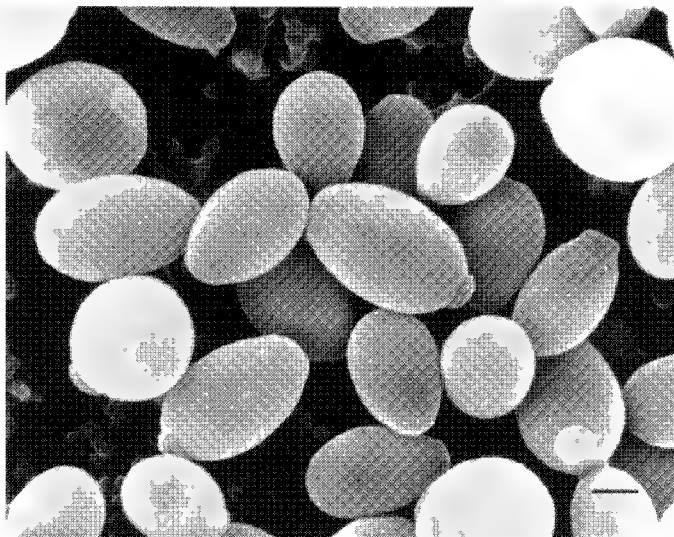


Fig. 1. Scanning electron micrograph showing cells of *S. cerevisiae*. Bud scars are visible at the ends of cells. Scale bar is 5 μm .

The yeast cell is surrounded by a mechanically refractory cell wall which may account for 20–30% of the cell solids (6). It surrounds the plasma membrane, which regulates the transport of chemical compounds into and out of the cell either by simple diffusion or by active transport. The cell wall consists of alkali-soluble β -glucan, alkali-insoluble glucan, and glucan- and mannan-containing glycoproteins. Linkages within glucans are generally β (1 $\rightarrow$ 3) and β (1 $\rightarrow$ 6). Some yeasts form viscous and sticky capsular material on the outside of the cell wall. Fimbriae (small, hairlike structures) may be present on the outside of the cell wall and may play a role in flocculation.

The small nucleus of the yeast cell is surrounded by a membrane or tonoplast, which has many pores with an average diameter of about 0.085 μm . In haploid cells of *Saccharomyces cerevisiae*, the nucleus contains 90% of the DNA of the cell, organized into 16 linear chromosomes ranging in size from about 150 to about 1600 kbp. The total DNA content is about 1.50×10^7 bp, or about three times the size of the *E. coli* genome (5×10^6 bp). The small average chromosome size makes it possible to separate intact chromosomes electrophoretically using a variety of pulsed-field electrophoresis systems (known as OFAGE, FICE, CHEF, and TAFE (7,8)), which have led to simple strain identification procedures and also allow rapid assignment of cloned genes to specific chromosomes (Fig. 2). Other yeasts may have their DNA organized into a smaller number of chromosomes: *Schizosaccharomyces pombe*, eg, possesses three chromosomes containing a total of about 1.4×10^7 bp.

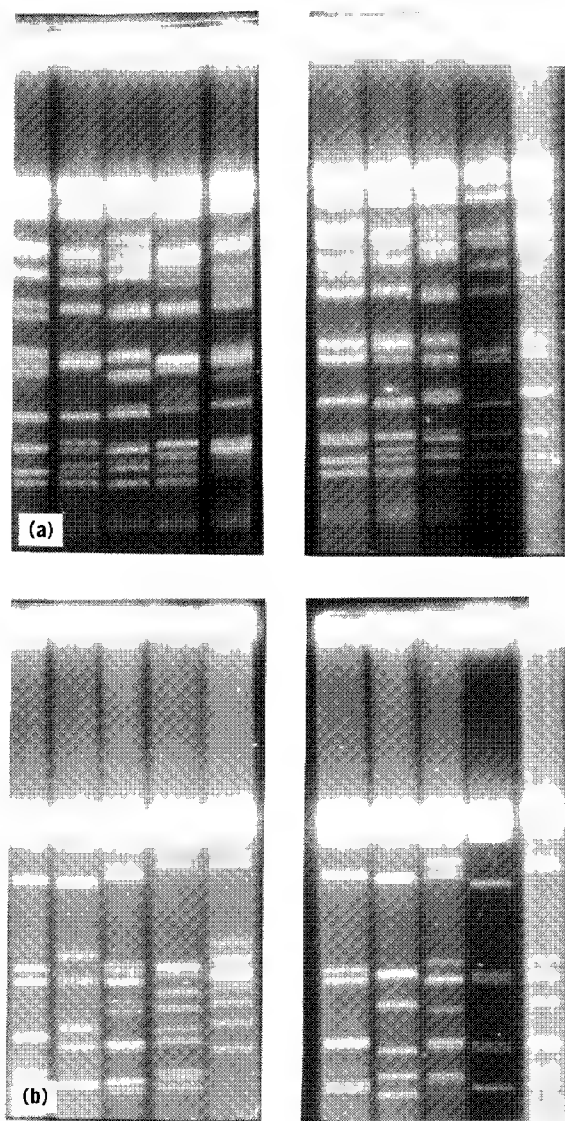


Fig. 2. Karyotype of 10 *S. cerevisiae* strains. Intact chromosomes have been separated electrophoretically by size in a TAFE gel. (a), gel run to separate most chromosomes. Large chromosomes are at the top, smaller at the bottom. Since most strains are polyploid, more than 16 bands may be observed. (b), Chromosomes from the same strains have been separated in a gel run to enhance resolution of the smaller chromosomes, corresponding to the 4–5 smallest bands in A.

Saccharomyces cerevisiae cells contain one or more mitochondria carrying about 5–10% of the cellular DNA, consisting of one or more copies of a single circular 14 kbp molecule. Mitochondria carry out the respiratory function of the cell and contain enzymes encoded by both nuclear and mitochondrial genes. Under anaerobic conditions, the mitochondria can degenerate, but will reform if the cells are aerated. The mitochondrial chromosome of *S. cerevisiae* is unstable, and about 1% of the cells in cultures of most strains contain defective mitochondrial genomes and are consequently incapable of respiration. Such cells give rise to small colonies on solid media ("petite," as opposed to the normal "grande," colony size). The largest organelle in the cell is the vacuole, which serves both hydrolytic and storage functions. Depending on culture and cell cycle conditions, one to four vacuoles may be found in a cell.

The cytoplasm of the cell contains most of the RNA, which may account for 7–12% of total cell solids, mostly as ribosomes. Lipid globules are also found in the cytoplasm, as well as carbohydrate storage materials in the form of glycogen and trehalose, which may account for up to 23% or more of cell solids, depending on culture and metabolic conditions.

Vegetative reproduction in yeast occurs mostly by budding or, in the instance of *Schizosaccharomyces* and *Endomycopsis*, by fission. The nuclear membrane remains intact. During budding, the nucleus of the mother cell divides and one portion enters the newly formed bud. The bud grows until it has reached a size almost as large as the mother cell and then separates from it, forming a bud scar on the surface of the mother cell and a birth scar on the surface of the bud. The number of bud scars can be counted, and the oldest cells of a population in stationary phase typically have 12–15 bud scars, apparently the maximum number of daughter cells that can be produced by a single mother cell. The genetic control of the cell division cycle is a major topic of yeast research and is not reviewed here. Fission in *Schizosaccharomyces* and *Endomycopsis* is by formation of a cross wall without constriction of the cell. Division at the cross wall or septum forms two separate cells.

Yeasts belonging to *Basidiomycetes* form external spores. In *Ascomycetes*, sexual reproduction occurs with the formation of spores in a cell that serves as a spore sac or ascus. Generally four, but sometimes eight or more, spores are formed in each ascus. This is sometimes reflected in the species name, ie, *Schizosaccharomyces octosporus* or *Kluyveromyces polysporus*. Sporulation in *Saccharomyces cerevisiae* is induced by starvation for nitrogen in the presence of a nonfermentable carbon source, eg, acetate. In principle, sexual reproduction in yeasts consists of an alternation between the haploid phase (one set of chromosomes) and the diploid phase (two sets of chromosomes). However, production strains of bakers' yeast (*S. cerevisiae*) are generally polyploid (often tetraploid, ie, containing four sets of chromosomes and producing diploid spores). Although sporulation in laboratory strains of bakers' yeast is quite efficient, with at least 70–80% of cells forming asci and most (90% or more) spores viable, production strains sporulate with low efficiency. At most, 10% of cells form asci and spore viability is generally 1% or less. The four ascospores of *S. cerevisiae* remain together within the wall of the mother cell; after removal of the wall by enzymatic digestion, they may be separated individually by micromanipulation and grown into colonies for genetic or physiological characterization. (Commercially available digestive enzymes include Zymolyase, isolated from *Arthrobacter luteus* and Glusulase, isolated from snail crop.) Other yeasts may exist predominantly in the haploid state, eg, *Saccharomyces rouxii*, in the diploid state as for *S. cerevisiae*, or in both states, eg, for *Saccharomycopsis lipolytica*. An alternation between the haploid and the diploid state may also occur without production of sexual spores (parasexuality).

Strain Improvement and Development

A variety of transformation techniques using *E. coli*-yeast shuttle vectors and yeast selectable markers, as well as efficient yeast promoters and signal

sequences, is generally available. Ref. 9 contains general reviews and descriptions of such procedures. For industrial strains, the rarity of dominant selectable markers for transformation sometimes poses a problem. Public disinclination to purchase genetically engineered strains of yeast for food use, eg, in baking, brewing, and wine making, has limited the use of such strains in most of the industry. Generally, work on strain improvement is, thus, carried out using the classical genetic approaches of hybridization and mutagenesis.

Hybridization. It is possible to isolate individual ascospores or vegetative cells of yeasts such as *Saccharomyces cerevisiae* with a micromanipulator. If individual cells of one strain are placed next to cells of another strain, conjugation or mating, may occur. The characteristic zygote formed is easily identifiable and may be isolated in turn. Such hybrids can also be produced by conjugation of spores of different species of the same genus (differences between species of *Saccharomyces* are usually minor, defined by carbon source utilization, and may not reflect large genetic differences). Ability to mate in *S. cerevisiae* is determined by a single locus called *MAT* (mating type) with two alleles. Cells expressing *MATa* are capable of mating with cells expressing *MATα*, and *vice versa*, but are not capable of producing ascospores. Cells expressing both mating types (*MATaα*), eg, produced by mating *MATa* and *MATα* cells, are not capable of mating but can produce ascospores. Strains of bakers' yeast are generally heterothallic: *MATa* and *MATα* strains display a stable mating phenotype. Many other strains, including some wine strains, are homothallic: cells of one mating type change to the other mating type after producing a bud and are capable of mating with their buds. The resulting zygote and its descendants are no longer capable of mating. Typically, only spore-spore matings are successful with homothallic strains.

Heterothallic and homothallic strains of *S. cerevisiae* differ at a single locus, *HO* (homothallism), which encodes a site-specific endonuclease. The change in mating type in homothallic strains begins when the endonuclease cleaves a site within the mating type locus, located on Chromosome III. This leads to the replacement of the resident mating type allele with a copy of the other; copies of both *MATa* and *MATα* are also present on Chromosome III, at loci known as *Hma* and *Hmα*, respectively. Only the allele present at *MAT*, however, is expressed.

Hybridization of incompatible strains is carried out by selecting rare matings or by protoplast fusions. Both procedures require the presence of selectable markers in both parent strains. Since dominant selectable markers in nuclear genes are hard to come by in industrial strains, use of mitochondrial antibiotic-resistance markers (eg, oligomycin or erythromycin resistance) is common. Rare matings are selected simply by mixing two appropriately marked parent strains, plating on nutrient agar and then replica plating to a selective medium on which neither parent strain can grow, but rare hybrids can. Confirmation of hybrid identity by karyotyping is generally required.

Protoplast fusion is carried out by digesting the cell walls of marked strains with hydrolytic enzymes (eg, Zymolyase) and mixing the resulting protoplasts in buffer to permit membrane fusion. Following plating on selective medium in an osmotically supportive top agar and cell wall regeneration, fusions can be recovered at low efficiency. Hybridizations between species of *Saccharomyces* and *Kluyveromyces* have been achieved, as well as between *Candida* and *Pichia* species.

Mutation. For industrial applications, mutations are induced by x-rays, uv irradiation or chemicals (nitrosoguanidine, EMS, MMS, etc). Mutant selections based on amino acid or nucleotide base analogue resistance or treatment with Nystatin or 2-deoxyglucose to select auxotrophs or temperature-sensitive mutations are easily carried out. Examples of useful mutants are strains of *Candida membranefaciens*, which produce L-threonine; *Hansenula anomala*, which produces tryptophan; or strains of *Candida lipolytica* that produce citric acid. An auxotrophic mutant of *S. cerevisiae* that requires leucine for growth has been produced for use in wine fermentations (see also WINE). This yeast produces only minimal quantities of isoamyl alcohol, a fusel oil fraction derived from leucine by the Ehdich reaction (10,11). A mutant strain of bakers' yeast with cold-sensitive metabolism shows increased stability and has been marketed in Japan for use in doughs stored in the refrigerator (12).

Fermentative and Respiratory Metabolism

Although most yeast species are strict aerobes, strongly fermenting yeasts such as *Saccharomyces cerevisiae* grow well under anaerobic conditions, but with a lower yield. The yield of yeast based on weight of sugar consumed is greater in the presence of oxygen, and oxygen decreases the rate of sugar consumption. Under aerobic conditions, alcohol is produced as the end product of metabolism if the sugar concentration exceeds 0.1–0.2% by weight. This is the reverse Pasteur or Crabtree effect and is also known as glucose inhibition or catabolite repression. In the presence of higher sugar concentrations, synthesis of respiratory enzymes such as cytochromes is inhibited.

The anaerobic pathways for glucose utilization by yeasts and other microbes lead to the formation of various alcohols and organic acids. Ethanol is the principal product of yeasts used in the beverage industry, formed from pyruvate via acetaldehyde by the enzyme alcohol dehydrogenase. Pyruvate itself is created in the Embden-Meyerhof-Parnass (EMP) pathway and the related hexose monophosphate (HMP) pathway. In either case, anaerobic fermentation results in 2 mol ethanol and 2 mol carbon dioxide from 1 mol glucose. This sequence also produces 2 mol adenosine triphosphate (ATP), supplying the energy for anaerobic growth. Common fermentable substrates are glucose, fructose, galactose, mannose, maltose, lactose, cellobiose, sucrose and some trisaccharides (see ETHANOL; FUELS FROM BIOMASS).

Respiratory, or oxidative, metabolism produces more energy than fermentation. Complete oxidation of one mol of glucose to carbon dioxide and water may produce up to 36 mol ATP in the tricarboxylic acid (TCA) cycle or related oxidative pathways. More substrates can be respired than fermented, including pentoses (eg, by *Candida* species), ethanol (eg, by *Saccharomyces*), methanol (eg, by *Hansenula* species), and alkanes (eg, by *Saccharomyces lipolytica*).

Composition, Nutrients, and Growth Rate

The elemental and vitamin compositions of some representative yeasts are listed in Table 1. The principal carbon and energy sources for yeasts are carbohydrates (usually sugars), alcohols, and organic acids, as well as a few other specific hydrocarbons. Nitrogen is usually supplied as ammonia, urea, amino acids or oligopeptides. The main essential mineral elements are phosphorus (supplied as phosphoric acid), and potassium, with smaller amounts of magnesium and trace amounts of copper, zinc, and iron. These requirements are characteristic of all yeasts. The vitamin requirements, however, differ among species. For laboratory and many industrial cultures, a commercial yeast extract contains all the required nutrients (see also MINERAL NUTRIENTS).

Table 1. Composition of Yeast^a

Component	Bakers' yeast	Brewers' yeast	<i>Candida</i> sp.
C, wt %	47.0		45.9
H, wt %	6.0		6.7
N, wt %	8.5		7.3
O, wt %	32.5		32.1
ash, wt %	6.0	6.4	7.8
Ca, wt %	0.06	0.13	0.57
Fe, wt %	0.003	0.01	0.01
Mg, wt %	0.13	0.23	0.13
P, wt %	1.0	1.4	1.7
K, wt %	2.0	1.7	1.9
Na, wt %	0.03	0.07	0.01
Co, mg kg		0.2	
Cu, mg kg	8.0	33.0	12.4
Mn, mg kg	5.9	5.7	38.7
Zn, mg kg	197	38.7	99.2
dry matter, wt %	94.0	93.0	93.0
crude fiber, wt %		3.0	2.0

ether extract, wt ⁰ o		1.1	2.5
protein (N $\times$ 6.25), wt ⁰ o	45	44.6	48.3
digestible protein, wt ⁰ o		38.4	41.5
thiamine, mg kg	90	91.7	6.2
riboflavin, mg kg	45	35	44
nicotinic acid, mg kg		448	500
pyridoxine, mg kg	40	43	30
biotin, mg kg	1.3		1.1
pantothenate, mg kg	65	110	83
folic acid, mg kg	15	10	23
choline, mg kg	4000	3885	2911

^a Refs. 13–15.

The specific growth constant for exponential growth, μ , is defined as $\mu \propto dt = dM / M$, where M is the yeast mass and t is time. Growth is also measured by either the generation time, ie, the time required for a doubling of the yeast population, or as the hourly growth rate. The relationship of these expressions is shown in Table 2. Exponential growth ceases when one or more nutrients become limiting. For aerobic cultures, the deficiency may be in oxygen, which is difficult to supply in large amounts because of its poor solubility.

Table 2. Relation of Generation Time to Hourly Growth Rate and Specific Growth Rate

Generation time, h	Hourly growth rate, C_t / C_0^a	Specific growth rate, μ , h ⁻¹
1	2.00	0.693
2	1.41	0.347
3	1.26	0.230
4	1.19	0.173
5	1.15	0.139

^a C_0 = concentration of yeast solids at time 0, C_t = concentration of yeast solids at time t (1 h).

Yeast-Fermented Foods and Beverages

Yeast is employed in the production of alcoholic beverages, baked goods, and yeast biomass and the production of fuel alcohol by fermentation (see Table 3). In the production of alcoholic beverages there is a 5- to 10-fold increase in yeast cells during fermentation. Bakers' yeast is supplied as a thick suspension (yeast cream, about 17% solids), as a moist press cake (about 30% solids), or as an active dry yeast (about 93–98% solids). There is little or no growth of yeasts during dough fermentation. Yeast biomass refers to the production of *Candida utilis* (torula) grown on sulfite waste liquors or ethanol and to smaller amounts of *Kluyveromyces fragilis* grown on cheese whey or whey permeate. These products are generally used as inactive, food-grade yeasts; some bakers' yeast is also used to produce food-grade, inactive dried yeast.

Table 3. Production of Yeast and Yeast-Fermented Foods, Beverages, and Alcohol

Product	Production, 10 ³ m ³	Yeast solids produced ^a , 10 ³ t
beer	22.9	44
wine	1.68	6
distilled beverages (50% ethanol)	0.36	4
baked goods (yeast raised)	11.3	90
fuel and industrial alcohol (100% ethanol)	3.0	67.5
yeast biomass (inactive)		10

^a Estimates based on production of 2.0–2.5 kg yeast solids per cubic meter of beverage at 8–12% ethanol.

Yeasts that have been grown in clear media, eg, clarified molasses or brewers' worts, can be recovered by centrifuging or filtration followed by pressing if desired. Yeasts that have been grown in distillers' mashes or grape musts, that contain insoluble particles, cannot be separated economically. All of the yeast produced in wine fermentations is either discarded as waste or spread with the lees or still residues as fertilizer. The excess yeast produced during the fermentation of distilled beverages and in fuel-ethanol production is recovered with the spent grains and sold as feed. Yeast produced during beer fermentations is largely combined with the spent grains and sold as feed as well; however, much is recovered separately by centrifugation and used for the production of either food-grade, inactive brewers' yeast or yeast extracts.

The fermentative activity of the yeasts is almost the same in all industries. Differences in fermentation times arise mainly from differences in temperature. For the baking industry, temperature is high, therefore fermentation time is short, with little time for cell multiplication and very high cell counts at the start.

Bakers' Yeast Production. Bakers' yeast is grown aerobically in fed-batch fermentors under conditions of carbohydrate limitation. This maximizes the yield of yeast biomass and minimizes the production of ethanol. Yeasts grown under these conditions have excellent dough leavening capability and perform much better in the bakery than yeast grown under anaerobic conditions.

Strains. All bakers' yeast strains are *Saccharomyces cerevisiae*. Historically they descend from distillers' yeast rather than brewers' or wine yeasts. Although many hybrid and mutant strains of bakers' yeast have been patented, few are in production. Pure cultures of yeast strains may be maintained on agar slants with frequent transfers, as frozen liquid stocks in 50% glycerol or 70% DMSO, or in lyophilized form. Suitable strains may be obtained from any culture collection. Additionally, since yeast sold to bakers and the general public is alive, it may be isolated from commercial cream, compressed, or dry yeast products. Strains differ in their ability to ferment in doughs containing different sugar levels. Some are best in "lean" doughs (ie, with no added sugar); some perform best with sugar additions of 6–8% of flour weight ("regular" doughs) or 20% of flour weight ("sweet" doughs). Yeasts used for the production of compressed or cream yeast (fresh yeast) in the United States, or for the so-called fast compressed yeasts of Europe, are ovoid with a cell diameter of 4–6 μ m. In contrast, yeasts used for the production of active or instant dry yeast are larger, ie, 5–8 μ m in diameter. Although dry yeast strains generally display lower leavening power on an equivalent solids basis when used as fresh yeast, they are more tolerant of drying conditions.

Raw Materials. Until the 1930s, grain worts were used as the principal carbon and energy sources for yeast production. Since then, cane and beet molasses have been used because they are much less expensive sources of fermentable sugars. More recently, increases in the cost of beet molasses and more efficient extraction of sugars from beet juice have caused yeast manufacturers to consider supplementing worts with additional carbohydrate sources such as corn syrup. Molasses contains about 50–55% fermentable sugar in the form of sucrose, glucose, and fructose. It is also a source of potassium and other minerals, and cane molasses is the principal source of biotin, a required nutrient for bakers' yeast production. The principal nitrogen

source is ammonia, added as ammonium sulfate, or as liquid, or anhydrous ammonia. Urea and many of the amino acids present in molasses are also sources of assimilable nitrogen. Since bakers' yeast contains ca 3% phosphorus (as P₂O₅), based on yeast solids, phosphate is added either as phosphoric acid or phosphate salts. Calcium and magnesium salts in smaller concentrations are also added. Iron, zinc, copper, manganese, and molybdenum are required as trace minerals. The vitamin requirements for biotin, inositol, and pantothenic acid are usually provided by molasses. Thiamin beyond that required for growth is usually added because it contributes to the fermentation activity of the final yeast.

Fermentation. Fermentations proceed from a series of smaller-scale fermentations, beginning with small flasks followed by transfer to larger and larger vessels (see ref. 16). There may be as many as eight stages in this sequence, producing perhaps 0.8 kg, 3.5 kg, 25 kg, 120 kg, 2.5 t, 15 t, and 100 t of yeast from a single colony isolated from a pure-culture slant. The earliest stages may be carried out under conditions approaching pure-culture fermentations, ie, with sterile media, sterile air, and presterilized vessels. For large-scale fermentations, molasses at 80° Brix is diluted to about 30° Brix and adjusted to pH 5. It is clarified by centrifugation to remove insoluble solids and sterilized by brief, high temperature treatment. The other nutrients are either included in the molasses wort or added separately during the fermentation. Thus ammonia is generally added separately because it is used both as a nutrient and as a means of pH adjustment. The fermentation is carried out at 30°C and pH 4.5–6.5; higher pH levels produce faster yeast growth but are conducive to the growth of contaminants as well.

The large-scale fermentations are carried out as fed-batch processes (formerly known as the Zulauf process). Medium is added to the seed yeast ("pitch") at a slow rate to maintain sugar concentrations in the fermentor at less than 0.1%. At higher sugar concentrations, excess alcohol forms, lowering the yeast yield and increasing the biological oxygen demand (BOD) of the effluent. Wort addition may follow a pre-set program (scheduled feed) or may be controlled in response to culture conditions, eg, ethanol levels (demand feed). With sufficient aeration, it is possible to greatly extend the exponential phase but in practice this is rarely done. Production cultures are normally harvested at yeast concentrations of about 17–20% (ie, 4–5% by weight yeast solids). In practice, the yield of yeast cell mass is 50% by weight of the sugar used.

Oxygen is a critical nutrient for yeast growth. One kilogram of oxygen is required for the growth of one kilogram of yeast solids. This oxygen is supplied by forcing air through a series of pipes with small holes, arranged near the bottom of the fermentor. The distribution of air and the formation of very small bubbles (to increase surface transfer area) are often aided by agitators. Oxygen is only sparingly soluble in water, and the saturation concentration at 30°C is only 7 ppm. During oxygen uptake by concentrated yeast suspensions, the oxygen level may decrease to only 5% of saturation. In large fermentors, the rate of air supply is about one volume of air per fermentor volume per minute (VVM). Alcohol is formed if the oxygen supply is limiting.

Finally, the rate of yeast growth must be limited to the specific growth rate constant of 0.23, again to limit alcohol formation. For an hourly specific growth rate of 0.23, the hourly increase in cell mass is 1.26 and the generation or doubling time is 3 hours. During yeast growth, heat is liberated and must be removed from the fermentor by internal cooling tubes or by pumping the fermentor contents through external heat exchangers.

The fermentation period is usually 12–18 h. For a 15-h fermentation, the generation time may increase from 3 to 5 and finally to 7 hours, allowing for roughly an eightfold increase in cell mass. Generally, aeration continues briefly after wort additions have been stopped to allow the yeast to mature. This results in a reduction in the number of budded cells as cells complete their last division cycle and increases storage stability of the final product. At the end of the growth period, the yeast solids concentration is about 5%. Experimentally, the solids concentration may be increased to 10% or higher, but in commercial practice oxygen demand is limiting.

Fermentations in larger vessels and the final trade fermentation are conducted under quasi-sterile conditions, and yeast growth is accompanied by some growth of contaminant bacteria. These are generally lactic acid-producing organisms but are sometimes coliform bacteria; the occurrence of *Salmonella* in fermentor liquids has not been reported. Massive contamination with *Oidium lactis* or wild yeasts has been reported.

Harvesting and Packaging of Yeast.

Cream Yeast. At a 5% solids concentration in the fermentor, the yeast occupies about 12% of the fermentor volume. It is harvested by centrifugation in nozzle centrifuges and washed several times with water. The final centrifugate is cooled and stored in refrigerated tanks. This yeast cream (so called because of its off-white color) may be sold directly in this form, since in large baking facilities, it may be piped directly to any desired location.

Compressed Yeast. A moist press cake is obtained by pressing yeast cream in plate-and-frame filter presses or filtration through a rotary vacuum filter. The press cake is mixed with emulsifiers (0.1–0.2 wt %) and its moisture content is adjusted to 70%. To form block yeast, the cake is then extruded through Teflon nozzles in the shape of thick strands with a rectangular cross-section and is typically cut into 0.45 kg cakes, which are wrapped in wax paper. Compressed yeast is also prepared as crumbled yeast in the form of irregularly shaped pieces packed in bulk containers. During pressing, mixing, and extruding, the temperature of the yeast cakes is 10–15°C. They must be cooled for about two days before shipment in refrigerated containers. Compressed yeast is a perishable commodity and must be refrigerated at all times. At a storage temperature of 5–8°C, there is a 3–5% loss of activity per week. Compressed yeast may be kept frozen for several months. However, after thawing, some browning and softening of the cake occur.

Compressed yeast is received by the baker in 0.45-kg cakes or as crumbled yeast in 22.7 kg bags and refrigerated before use. Deliveries may be on alternate days or once a week, and the stock must be rotated so that the yeast is never kept longer than 1–2 weeks in the bakery. The yeast cakes may be mixed directly into the dough by being placed directly on top of the flour in the mixer in a straight dough fermentation. More commonly, crumbled yeast (or cream yeast) is used in the preparation of preferments. The crumbled yeast is slurried with a small portion of the water and other ingredients (oxidants, enrichment tablets, and yeast nutrients) in a premix tank, allowed to ferment and then pumped into the mixer.

Compressed yeast is also sold in supermarkets in 18-g and 56-g packages. Since this product contains approximately 10% added starch to increase its shelf life, it has a lower protein content and fermentative activity than the compressed yeast sold to bakeries.

Active Dry Yeast (ADY). The production of active dry yeast is very similar to the production of compressed yeast. However, a different strain of yeast is used and the nitrogen content is reduced to 7% of solids compared with 8–9% for compressed yeast. The press cake made with the active dry yeast strain is extruded through a perforated plate in the form of thin strands with a diameter of 2–3 mm and a length of 3–10 mm. The strands are dried on endless belts of steel mesh in drying chambers (a continuous process) or in roto-louvre dryers (a batch process), with the temperature kept below 40°C. Drying time in drying chambers is 3–4 h and in roto-louvre dryers is 6 h or more. The final moisture level attained is 7.5–8%.

Where storage stability and convenience are important considerations, ADY has largely replaced compressed yeast. The yeast is packaged in hermetically sealed aluminum foil pouches for domestic use, in 0.45–0.91 kg cans for institutional use, or in 10-kg foil pouches. Since a minimum storage life of 1 year is guaranteed, the yeast is packed in a nitrogen atmosphere or under vacuum, and the integrity of the package is at least as important as the original activity of the yeast. For use in wholesale bakeries, active dry yeast is usually delivered in polyethylene-lined drums and is not protected by an inert atmosphere; accordingly, it should be used within 2–3 months. The yeast requires separate hydration in water, preferably at 35–40°C before addition to other dough ingredients. A period of 5–10 minutes is sufficient for full rehydration and optimum activity.

Instant Active Dry Yeast. Instant ADY (IADY or HADY) production is similar to ADY production but requires a different strain of yeast. After pressing, the yeast is extruded into noodles 0.2–0.5 mm in diameter and 1–2 cm long and deposited on a metal screen or perforated plate in a fluid-bed air dryer. Drying time is shorter than with ADY, about 1–2 hours in practice, with a final moisture level of 4–6%. Instant active dry yeast does not require separate rehydration. It is always packaged in a protective atmosphere or under vacuum. On an equivalent solids basis, the activity of IADY is greater than that of regular ADY, but still less than that of compressed yeast.

All active dry yeasts make doughs softer, more extensible and less elastic. This slackening is more pronounced with the regular ADY than with IADY. The effect is believed to be due to the leaching of reducing agents, especially glutathione, from the dried yeast during rehydration, before the cells have fully recovered their activity. Up to 50 ppm (yeast solids basis) of glutathione can be leached from frozen yeast (17), although this may not be enough by itself to have a substantial effect. Slackening can be counteracted with higher levels of oxidants in the dough.

Yeast in Baked Goods. Baked goods can be leavened with yeast, chemicals, the foam of egg white, or steam. There is a preference for yeast-leavened baked goods because of their desirable flavor. Therefore, only those products are chemically leavened in which yeast does not perform well, generally as a result of high osmotic pressure. The osmotic pressure of doughs depends inversely on moisture content and directly on the concentration of salt and sugars. Cookies and highly sweetened cake doughs have an osmotic pressure too high for adequate fermentation by yeasts. Although some attempts have been made to grow yeasts with exceptional tolerance to high osmotic pressure, none has been used commercially. There are some products

in which either yeast or chemical leavening, or both, are used, eg, doughnuts and pizza dough.

Another requirement for proper leavening is the presence of a protein matrix sufficiently elastic to trap small carbon dioxide bubbles. Wheat gluten fulfills this requirement. Rye protein is less suitable and the proteins of other cereals, eg, rice, oats, or corn, are practically useless.

White Pan Bread. More than 50% of all the flour used in baked food production is made into bread, rolls, and buns, and white bread is the predominant dough type. The basic method of production is the sponge-and-dough method. The sponge generally contains 60–70% of the flour, all of the yeast, and yeast food. It undergoes a 3–4 h fermentation, when the sponge, the remainder of the flour, and all other ingredients (salt, sugars, fat, etc) are mixed into a dough and permitted to rest for 20–30 min. The dough is divided, rounded, molded, and panned. The pans are placed in a proofing cabinet where the final fermentation takes place before baking. Mixing of the sponge, the sponge fermentation and the mixing of the dough are batch processes. From the dough dividing step to the final baking, the process is generally continuous.

Since the 1950s and 1960s other production methods have been used. For example, the sponge can be replaced by a liquid sponge or preferment, which contains little or no flour but most or all of the water, yeast, yeast nutrients, and a suitable buffer. Addition of the buffer prevents the pH from dropping below 4.5. The preferment must remain sufficiently fluid that it can be pumped. The temperature is generally set at 25–30°C and fermentation time is 2–4 h. Another method is the use of continuous dough mixing, which must be combined with the use of liquid preferments. Continuous high-speed mixing produces a very extensible, slack dough with a uniform distribution of fine bubbles of carbon dioxide. The resulting bread has a very fine grain and is very soft.

All the preceding methods have the advantage of flexibility. Depending on the course of the sponge fermentation, alterations can be made in dough mixing time or in the addition of other ingredients, eg, oxidants, to the dough. In case of equipment failure, sponges can be held for several hours without spoiling, and preferments can be cooled and stored overnight. Wholesale bakers also use straight dough methods, in which all of the ingredients of a bread formula and the water are mixed in a single operation. The dough is then fermented for 2–4 h. The operations following fermentation are then the same as those for sponge and dough methods. More detailed descriptions of actual bakery methods are found in Reference 18.

The function of yeast in bread making is threefold: it serves as a leavening agent, it contributes to flavor, and it matures the dough. In contrast to chemical leavening, which is a rapid reaction, yeast leavening is slow and proceeds as long as fermentable sugar is present. Flour contains about 1% fermentable sugar, mostly glucose and maltose, which is generally fermented within the first 1–1.5 h. However, flour also contains damaged flour granules (6–12% by weight), which are hydrolyzed by amylases to maltose. A β -amylase is present in the flour, and α -amylase derived from barley malt or fungi may be added as well. In doughs without added sugar, glucose is fermented before fructose; maltose fermentation is initiated only after glucose is essentially exhausted, and continues as long as maltose is produced from starch.

In practice, most bread doughs contain 4–8% (based on flour weight) added sugar, usually sucrose, glucose, corn syrup, or high-fructose corn syrup. Most of these sugars are consumed during fermentation, but some residual sugar appears in the bread. With the addition of 6.7% sucrose, for example, the residual sugars in the baked bread are fructose (2.6%), glucose (1.7%), and maltose (0.7%). Lactose (milk sugar) is not fermented by bakers' yeast, and addition of milk solids or whey results produces residual lactose in the bread.

The pH of doughs or preferments has little effect on yeast fermentation rates unless the pH drops below 4.5. Prior to fermentation, the pH of a dough is 5–6. During fermentation it drops to about 4.8 due to production of carbonate (from dissolved carbon dioxide) and organic acids. During baking there is a rise in pH as carbon dioxide is driven off. Temperature has a considerable effect on yeast activity: up to 38°C, activity increases but diminishes at higher temperatures. At 52°C, bakers' yeast cells die within 10 min. Bakery sponges, straight doughs, and preferments are generally set at 24–26°C. During fermentation there is a rise of 2–5°C resulting from the metabolic activity of the yeast. In the case of continuously mixed doughs, the temperature may reach 35°C outside the mixer. Proof box temperatures are generally 30–40°C. During baking, the dough temperature rises rapidly and yeast activity ceases after about 10 min. During this time, expansion of the dough in the pan (the so-called oven spring) occurs, resulting from continued fermentation, reduced solubility of carbon dioxide in the dough water, thermal expansion of carbon dioxide bubbles in the dough water, and evaporation of ethanol and volatile organic acids.

The effect of osmotic pressure on yeast activity is of great importance, and is often overlooked. At salt concentrations up to 1.5%, the effect is slight; salt concentrations of 2–2.5%, which are common in bread doughs, inhibit yeast activity considerably. Likewise, sugar concentrations above 4% produce apparent inhibition. Consequently, yeast-raised sweet doughs (15–20% sugar), contain very high yeast concentrations.

Ethanol inhibits yeast fermentation. Although dough alcohol concentrations rarely exceed 3% and are not inhibitory, this may be important in wine fermentations where ethanol concentrations may reach 11–12%. Mold inhibitors are commonly added to bread doughs. These may be calcium propionate (0.25–0.50% of the flour) or small amounts of vinegar. These mold inhibitors may inhibit fermentation to a slight degree unless the yeast is specially treated by the manufacturer during growth.

There is little or no growth of yeast during dough fermentation. However, budding can be observed during sponge fermentation or during the proof stage of a straight dough fermentation. The addition of yeast nutrients (about 0.05% ammonium sulfate or ammonium chloride) is common.

The effectiveness of yeast leavening depends on the development of a protein matrix to retain carbon dioxide. The matrix can be obtained with wheat flours, which contain 10–14% protein. A portion of the carbon dioxide produced during the fermentation escapes from the dough and is not available for leavening. This is important during proofing when the volume of the bread is determined. The amount of high speed mixing is critical to development of the elastic protein matrix. Undermixed or overmixed doughs do not retain gas well. The protein matrix does not develop with corn flour and consequently corn breads have a coarse, grainy crumb.

Yeast fermentation is also central to flavor development. Flavor is derived from fermentation by-products, eg, higher alcohols, esters, aldehydes, and other carbonyl compounds. However, in contrast to alcoholic beverages, the flavor of baked bread is produced by the interaction of these fermentation by-products with other dough constituents, especially during baking. A considerable part of bread flavor is formed in the crust during the browning. After baking and cooling, this flavor diffuses into the bread crumb. Despite the identification of more than 100 flavor compounds, very little is known about these reactions.

The role of yeast in fermenting dough maturation is even less clear. The alcohol and carbon dioxide developed during fermentation must influence the elastic properties of the protein matrix. However, experimental procedures that would permit this to be checked in the absence of yeast have not been developed.

Sourdoughs. The microflora of flour includes large numbers of lactic acid producing bacteria. In the past, bread dough fermentations were mixed yeast and bacterial fermentations; this is still true of breads made exclusively from rye dough (in the United States, rye bread generally contains only small amounts of rye flour mixed with large amounts of wheat flour). Sour rye doughs and wheat doughs are inoculated with bacterial starter cultures as well as yeast. The lactic acid bacteria normally used in sourdough fermentations are *Lactobacillus plantarum*, *L. brevis*, or *L. fermenti*. The production of organic acids by these bacteria causes a much larger drop in the pH of the dough than occurs in a fermentation using yeast only.

The production of San Francisco sourdough bread is a typical sponge-and-dough fermentation, and a portion of each sponge is retained for inoculation of the succeeding sponge. The yeast used in this fermentation is *Saccharomyces exiguus* (19); the bacterium is *Lactobacillus sanfrancisco* (20). The same yeast is also responsible for the fermentation of the Italian panettone, a traditional Christmas fruitcake.

The production of soda crackers is also based on a mixed fermentation. Doughs for cracker production are inoculated with very small amounts of bakers' yeast. During the first 3–5 h of the 18-h fermentation, yeast activity predominates; thereafter bacterial fermentation causes a rapid decrease in pH through formation of lactic acid.

Brewing. Beer brewing has been traced back to ancient Egypt (at least to about 1000 BC), in a form related to modern practice. Recent findings suggest that sprouted barley, dried, cooked, and ground, along with crushed uncooked wheat or malt was treated with hot water (21). The filtered extract was then inoculated with yeast. Because the extraction was inefficient, the same mixture of grains may have been extracted repeatedly. It is not clear whether the beer was flavored, as modern beers are by the addition of hops.

The basic raw materials for the production of beer are sweet worts formed by enzymatic hydrolysis of cereal starches. The principal cereal is barley which, after malting, is also the source of enzymes that hydrolyze starches, glucans, and proteins. In some countries, eg, Germany, the mash bill consists

entirely of malted barley. In other countries, adjuncts such as corn, rice, corn syrup, or glucose are common. In the United States, the mash bill typically contains 60 wt % barley malt and 40 wt % corn or rice. In the UK, sugar syrups are often used.

Malt is produced by steeping barley in water for several days. This initiates germination, which activates several enzymes important in the digestive process that occurs in malt. The germinated barley is then dried and kilned at about 80°C, or higher for darker malts.

Brewers' wort is made by mashing the barley malt, or malt plus adjuncts, with water. In the infusion method, the temperature of the mash is gradually raised to 60–65°C through a number of steps. Heating is gradual, and the mash is maintained for some time at about 52°C to permit the proteolytic enzymes to act. At the final temperature, amylases catalyze the hydrolysis of starches to maltose, oligosaccharides, and dextrins. Adjuncts are generally heated to boiling to gelatinize their starches. The boiled adjuncts are added slowly to the malt mash to raise its temperature in the desired sequence. Thus, they contain no essential enzymes that must be preserved for mashing; their starch is readily hydrolyzed by the amylases of the malt mash. Finally, the mash is filtered to separate the insoluble constituents such as barley hulls from the soluble constituents. The filtrate or sweet wort is then pumped into the brew kettles where it is boiled with hops. This inactivates the enzymes, extracts the bitter hops flavor, and sterilizes the wort.

Brewery fermentations are carried out with pure-culture yeasts. Stock cultures of brewers' yeasts may be maintained on agar-containing solid media containing malt extract, glucose and peptone, or in a lyophilized form. During the aseptic propagation of the seed yeast, there is a ten-fold increase in cells from one step to the next. Often, yeast from a preceding commercial beer fermentation is inoculated into one or several (up to 10) succeeding fermentations. The fermentations are not carried out under strictly aerobic conditions. Although some sterile air is forced through the wort, the yeasts' metabolism is not respiratory because of the high concentration of sugar.

Two pure-culture yeasts, top- and bottom-fermenting, are used in brewing. Although both are biologically members of the species *S. cerevisiae*, bottom-fermenting yeasts are referred to as *S. uvarum* or *S. carlsbergensis*. Top-fermenting *S. cerevisiae* is used for the production of ale and *S. uvarum* for the production of lagers. The yeasts are indistinguishable microscopically. The presence of melibiase (α -galactosidase), which permits complete fermentation of the trisaccharide raffinose, characterizes *S. uvarum*. *S. cerevisiae* ferments only the fructose portion. This diagnostic feature is of no practical importance for brewing, since worts have little or no raffinose. In general, top-fermenting yeasts have stronger respiratory systems than bottom-fermenting yeasts and are often less flocculant.

The suitability of brewers' yeast strains depends largely on the rate and extent of growth, rate, and extent of fermentation, degree of flocculence, and influence on the flavor and aroma of the beer. Fermentation rates of brewers' yeasts are of the same order of magnitude as wine, bakers', distillers' or sake yeasts. Typical activities are 600 and 700 mL of CO₂ per hour per gram yeast solids in glucose and maltose medium at 25°C (22), and 265 mL h⁻¹ g at 25°C in wort (23). The concentration of fermentable sugars in worts varies with the concentration of the wort and with the composition of the mash. About 90–92% of the wort solids are carbohydrates, of which 70–80% are fermentable. As a rule, a barley wort contains 0.5–1.0% glucose, 3.5–6% maltose, 1.1–1.8% maltotriose, and 0.1–0.5% sucrose and fructose (24). The transport of glucose, sucrose, and fructose into the yeast and their rate of fermentation are similar. However, the fermentation of maltose undergoes adaptation and de-adaptation (although some strains are constitutive for maltose fermentation). The rate-limiting step appears to be transport of maltose across the membrane. The synthesis of the maltose transport system is repressed by glucose in the medium. Adaptation begins when the concentration of glucose in the wort decreases below 0.6%. Maltotriose fermentation takes place when the concentration of maltose falls to a low level; some strains do not ferment maltotriose and consequently produce beers with a high concentration of the trisaccharide. Maltotetraose, isomaltose, and higher glucose polymers are not fermented.

Fermentation. Lager beer fermentations with the bottom-fermenting *S. uvarum* are carried out at 10–16°C for 6–10 days. Fermentation is followed by a lager period of up to several months at 2–5°C. Ale fermentations with the top-fermenting *S. cerevisiae* are carried out at 15–22°C for 3–5 days. The rate of fermentation is highly dependent on temperature. For example, at 25°C, the rate of fermentation is about three to four times faster than at 10°C. The pH drops from about 5.5 to about 4.3–4.5 during the course of fermentation. The rate of maltose fermentation is almost constant at pH 3.5–5.0 (although it drops sharply outside this range) and the optimum pH for maltotriose fermentation is 4.3–5.4. Thus, the rate of wort fermentation is not greatly affected by this drop. Fermentation and yeast growth are inhibited by high concentrations of ethanol and dissolved carbon dioxide. At ethanol concentrations of 5–6% and carbon dioxide levels of 0.4–0.5%, the inhibition is minimal. Since the alcohol content of U.S. lager beers ranges from 3.4–3.8%, and only reaches about 8.7% in British stout, any inhibition is slight. Initially, brewers' fermentations contain 5–10 $\times 10^6$ yeast cells per mL, corresponding to about 0.75 g yeast solids per liter. Fermentation results in a four- to eight-fold increase in cell mass. The basic carbon source for this growth is carbohydrate. Some oxygen is available at the start of the fermentation, by aeration of the wort. Very little oxygen, however, is required for fermentative yeast growth. Nitrogen is supplied by the amino acids in the wort. Some amino acids, particularly asparagine, serine, threonine and lysine, are assimilated rapidly, while others are absorbed more slowly. Proline is not absorbed, and the resulting beer contains a relatively high concentration of this amino acid. Brewers' worts usually contain sufficient minerals and vitamins to support yeast growth and fermentation.

Flocculation. The exact degree of flocculence is very important for both top- and bottom-fermenting yeasts. If the yeast is too flocculent, a large portion rapidly settles out, delaying fermentation. If the yeast is too powdery, the fermentation is rapid but the yeast does not settle out at the end of fermentation. Genes involved in flocculation have been identified; physical and chemical factors such as divalent cation (Ca and Mg) concentration, pH, temperature, and net cell surface charge also play a role. In some strains, flocculation is apparently affected by hairlike appendages, fimbriae, on the cell wall.

Flavor and Aroma. The flavor compounds, fermentation by-products, are higher alcohols (known as fusel oils), esters, diketones, aldehydes, organic acids, and sulfur compounds. The particular yeast strain used affects the formation of higher alcohols and esters. Although the flavor compounds can be identified analytically, it is not always possible to specify their effect on flavor. Thus, the higher alcohols are individually present in concentrations below their threshold of perception, but the combination affects flavor. Their total concentration may be 50–100 ppm; about 50–66% of this consists of isoamyl alcohol, isobutanol, and propanol. Amyl and butyl alcohol are negligible. The concentration of esters in U.S. lagers is 25–50 ppm. Ethyl, isoamyl, and phenethyl acetates contribute a fruity flavor. Ethyl acetate forms from the reaction of coenzyme A with ethanol at a rate greatly dependent on temperature. Maximum formation occurs at 20–25°C, and top-fermented ales contain large amounts of this ester.

Diacetyl, acetoin, and diketones form during fermentation. Diacetyl has a pronounced effect on flavor, with a threshold of perception of 0.1–0.2 ppm; at 0.45 ppm it produces a cheesy flavor. U.S. lager beer has a very mild flavor and generally has lower concentrations of diacetyl than ale. Diacetyl probably forms from the decarboxylation of α -ethyl acetolactate to acetoin and consequent oxidation of acetoin to diacetyl. The yeast enzyme diacetyl reductase can irreversibly reduce diacetyl to acetoin. Aldehyde concentrations are usually 10–20 ppm. Their effects on flavor must be minor, since the perception threshold is about 25 ppm.

Organic acids, including carbon dioxide, lower the wort pH during fermentation. The principal acids formed are lactic, pyruvic citric, malic, and acetic acids, at concentrations ranging from 100–200 ppm. The main sulfur compounds formed during fermentation and their perception thresholds are as follows: H₂S (5–10 ppb); ethanethiol (5–10 ppb); dimethyl sulfoxide (35–60 ppb); and diethyl sulfide (3–30 ppb). At low levels, these may have a desirable flavor effect; at higher levels they are extremely undesirable. Sulfur dioxide also forms during fermentation, at concentrations of 5–50 ppm; its presence can be tasted at levels above 50 ppm.

Beer taste can be spoiled by contaminating bacteria or yeasts. The most common bacteria are lactic and acetic acid producers and *Zymomonas*. Wild yeasts can be anything other than the intended strain: *S. uvarum* is considered a contaminant of ale fermentations and *S. cerevisiae* a contaminant of lager fermentations. The common wild yeast contaminants are *S. diastolicus* and species of *Pichia*, *Candida*, and *Brettanomyces*. It may be noted that the flavor of beer may be improved by the ability of yeast to adsorb bitter substances extracted from hops, such as humulones and isohumulones.

Wine. The earliest known wines were made in Iran about 5400–5000 BC (25). The species of grape used is unknown and may have been either the wild grape *Vitis viniferus sylvestris* or a cultivated precursor of the modern wine grape *V. viniferus viniferus*. The source of the yeast used, and the procedures used are completely unknown. In modern times, grapes (about 21–23% sugar) are pressed; the liquid must is either separated and allowed to settle for 1–2 days (for white wines) before inoculation with yeast, or the whole mass is directly inoculated with yeast (for red wines). In either case, while the initial fermentation takes place, the carbon dioxide formed by fermentation excludes air and prevents oxidation. White wines are transferred to a second fermentor (racked) near the end of fermentation and kept isolated from the air while solids, including yeast, settle out, a process that requires about six

weeks. Additional rackings, filtration or centrifugation may be required to clarify the wine and remove tartrates prior to bottling.

Red wines are fermented in the presence of the grape skins to extract pigments and tannins, until 70–90° of the sugar has been fermented, and then pressed. The juice undergoes a residual fermentation to consume the remaining sugar. At this point the wine is racked to separate the yeast; a second, bacterial malolactic fermentation, which transforms malic acid to lactic acid, occurs at this stage and reduces the acidity of the wine. Red wines may be aged in wooden casks to improve flavor and precipitate tannins and tartrates. Additional filtrations or centrifugations may be employed before bottling to clarify the wine.

Addition of up to 200 ppm sulfur dioxide to grape musts is customary. Strains of *S. cerevisiae* and *S. bayanus*, grown in the presence of sulfite, become tolerant of fairly high concentrations of SO₂. Cultures propagated in the winery are added in liquid suspension, usually at 1–2° of the must volume. Many strains are available in pure culture. Factors such as flocculence, lack of foaming, fast fermentation, lack of H₂S and SO₂ formation, resistance to sulfur dioxide and other inhibitors, and flavor production will affect strain choice. No strain possesses all the desired properties.

Pure culture yeasts are available on slants or in lyophilized form. They may be propagated in wineries without complete asepsis (26,27), since the low pH of grape must inhibits many bacteria; wine fermentations are mixed cultures naturally. Production of active dry wine yeasts (WADY) began in the 1960s. These are produced by bakers' yeast companies and are grown by methods resembling those used for bakers' yeast production, described earlier. The dry yeasts have excellent storage stability, up to a year or more if packaged under an inert atmosphere (N₂, CO₂, or vacuum). First introduced into the United States and then Australia, they are now being introduced into European winemaking as well. A number of strains of *S. cerevisiae*, *S. bayanus*, and *S. fermentati* are available.

In addition to strains of *S. cerevisiae*, the term wine yeast refers to genera that occur naturally in grape musts and participate in spontaneous fermentation in wine. The number of yeast cells on intact young grape berries is not large. However, it increases as the grapes mature and there are many on injured grapes leaking juice. Essentially none of these are *S. cerevisiae*, and most yeasts that take part in spontaneous wine fermentations derive from cells adhering to wine-making equipment such as crushers and presses. Generally, the early fermentation is dominated by *Kloeckera apiculata*, *Metschnikova pulcherrima*, and *Torulopsis stellata*. *S. rosei* appears between the start and the main fermentation, which is dominated by *S. cerevisiae* and *S. uvarum*. *S. cerevisiae* and *S. bayanus* dominate the final stages. The succession of species described here depends largely on their alcohol tolerance, less tolerant species disappear as the alcohol content increases.

Spontaneous fermentations are used for wine production in France, some other European countries and in South America. In recent years, smaller California wineries have begun experimentation with spontaneous fermentations as well. They generally start more slowly than fermentations inoculated with commercial dried yeast, are more difficult to control, and may suffer from growth of undesirable contaminants. However, it is claimed that the resulting wines possess better organoleptic properties, particularly more complex flavor and aroma.

A number of different species dominate the fermentation at different times; *Saccharomyces* is dominant only towards the end. *K. apiculata*, with cell dimensions of (2–4.5) μm × (5–8) μm, usually dominates the early stages, along with *Metschnikova pulcherrima* and *Torulopsis stellata*. It stops fermenting and growing at ethanol concentrations of 4–5°. Cells of *K. corticus* and *K. africana* may also be observed. *Kloeckera* is the imperfect form of *Hanseniaspora* and does not form ascospores. The small cells are pointed or lemon-shaped and show bilateral budding. *Metschnikova pulcherrima* (*Candida pulcherrima*) propagates by multilateral budding and forms large asci with needle-shaped ascospores. Growth is inhibited at 4° alcohol as well. *T. stellata* (*T. bacillaris*) is an imperfect fungus. Cells are ellipsoidal and propagate by multilateral budding. Its alcohol-producing ability varies. Of the *Pichia* yeasts, the predominant is *P. membranefaciens*, a film-forming yeast. It forms a continuous film on the surface of young wines. Strictly aerobic, it does not produce ethanol, but tolerates high concentrations. The cells are ellipsoid to cylindrical, (2.5–4.5) μm × (5–14) μm. Vegetative reproduction is by multilateral budding; 2–4 sexual spores are formed. *Hansenula* yeasts propagate vegetatively by multilateral budding and form 2–4 hat- or saturn-shaped ascospores. Cells are (2.5–5) μm × (5.5–20) μm. Some species may produce up to 10° ethanol as well as ethyl acetate and other esters. *Candida krusei*, *C. vini*, and *C. valida* are film-forming wine spoilage yeasts. *Saccharomyces ludwigii*, the only species in its genus, is another wine spoilage yeast and is highly resistant to sulfur dioxide. The cells are lemon-shaped, (5–8) μm × (10–30) μm. Four spherical ascospores may be formed. *Schizosaccharomyces* species are spore-forming yeasts that divide by fission rather than budding. *S. pombe* is known for its ability to degrade malic acid to ethanol and carbon dioxide under anaerobic conditions. Although not often found in spontaneous fermentations, it has been added to fermentations as a pure-culture yeast. *Brettanomyces* cells are ovoid or pointed at one end. *B. intermedius* is a wine-spoilage yeast. During fermentation a strong characteristic odor is observed, and under aerobic conditions, large concentrations of acetic acid are produced. It can tolerate up to 12° ethanol, but is very sensitive to sulfur dioxide.

Saccharomyces yeasts are rapid fermentors. *S. cerevisiae* and *S. bayanus* produce up to 18–20° ethanol. The cells are ovoid to spherical, elliptical, or elongated (especially under conditions of nitrogen starvation). Vegetative propagation is by multilateral budding. *S. uvarum* and *S. rosei* occur earlier in the fermentation, when *S. rosei* may produce up to 6–8° ethanol before being overgrown by the other *Saccharomyces* yeasts. *S. cerevisiae* may produce up to 18–20° ethanol (28).

Distilled Beverages. Distilled alcoholic beverages are made by fermenting sugars from grains or fruits followed by distillation. The common cereal grains used for whiskey are barley, wheat, corn, or rye. Vodka is made from potatoes, rum from molasses, and brandy from grape or other fruits. Grain or potato starch must be converted to a fermentable form before use. Except for some Scotch or Irish whiskeys, *whole mashes* of grains are fermented. That is, fermentation and conversion of starch to fermentable sugar proceed simultaneously. The cereal grains, except for barley malt, are cooked in water to gelatinize the starch. The water temperature, in pressurized vessels, may exceed 100°C. After cooling the mash to about 65°C, a slurry of malted barley is added, to about 5–10° of the total mash. Amylases in the malt convert the starch to fermentable sugars. (Although fungal amylase is sometimes used for this purpose, the barley malt also provides some flavor to the whiskey.) The mash is cooled to 25–25°C before inoculation with yeast.

Distillers' yeasts are specialized strains. Many distillers maintain their own proprietary yeasts or have them grown by commercial yeast producers. Commercial bakers' yeasts may be used for the production of neutral grain spirits. The yeast is propagated on cooled mash before inoculation into the final mash at 3–5 vol °. Sometimes the inoculum is a portion of an actively fermenting mash ("backstocking"). In the UK, residual brewers' yeast is sometimes used. Distillers' fermentations are generally carried out at pH 4.8–5.0, and at an initial temperature sufficiently low that the final temperature does not exceed 30°C. Fermentation time is 2–5 days; the latter stages are very slow because after 30 h the limiting factor is starch conversion (29). Alcohol concentrations at the end of fermentation are about 7–9°.

Lactic acid bacteria are common contaminants of distillers' fermentations. *L. lactis* may produce excessive amounts of volatile acids. Some species convert glycerol to *β-proprionaldehyde*, which may break down to acrolein during distillation, producing an acid odor.

Fermentation of fruit juices other than grapes, principally apple juice, is the same as that in the production of grape wine. For the production of rum, cane molasses is diluted to a sugar concentration of 15–29°; the sucrose, glucose, and fructose is completely fermented within 36 h at 30–32°C.

Sake. Production of Japanese rice wine begins with the preparation of a culture of *Aspergillus oryzae* (30) by inoculating steamed polished rice. The mold produces extracellular amylases and proteases that act on the rice, liberating glucose and amino acids. This *koji* is inoculated into a mixture of water and additional steamed rice that is allowed to ferment for a period of about 10 days (the *moto*). At this time, it may be inoculated with a culture of sake yeast, *Saccharomyces sake*, a specialized subspecies of *S. cerevisiae*. During this initial period, continued growth of *Aspergillus* and the action of lactobacilli produce a solution high in glucose (26–28°) and organic acids, which favors the growth of the sake yeast. After approximately two weeks of yeast fermentation, there are three successive additions of *koji*, steamed rice, and water to the *moto*. A final fermentation of three weeks following the last addition produces fresh sake, with an alcohol content of 18–21° and a significant accumulation of succinic, lactic, and malic acids. This is filtered, held one month to allow settling of the remaining solids, pasteurized, and bottled. It is frequently diluted to an alcohol content of 15–16°.

Although *S. sake* is biologically a strain of *S. cerevisiae*, it differs in several important aspects from Western strains. Sake yeasts will grow at relatively high level of lactic acid (over 2°), and excrete succinic acid while fermenting. They continue to ferment at alcohol levels much higher than tolerated by other yeasts and will grow, albeit very slowly, at the very high osmotic pressures found early in the *moto* stage; some strains will continue to grow at sugar concentrations of 45° (31).

Soy Sauce (Shoyu). Production of soy sauce is a two-step process in which a mixture of boiled soybeans and crushed roasted wheat is first

inoculated with 1–2° of a culture of *Aspergillis oryzae* or *A. soyae* previously grown on polished rice (32). The mixture is allowed to ferment 2–3 days at 30°C and then transferred to large vessels along with an approximately equal volume of 18–19° brine. The subsequent fermentation, which traditionally lasts 2–3 years, is first dominated by lactic acid-producing bacteria including species of *Pediococcus*, *Bacillus*, and *Lactobacillus*. As the pH of the fermenting mash (*moromi*) drops owing to acid produced by the bacteria, a yeast fermentation begins, dominated by strains of *Saccharomyces rouxii* tolerant of high salt and osmotic pressure. Proper flavor development, however, requires the presence of species of *Torulopsis* as well.

The function of the *Aspergillis* fermentation appears to be primarily the breakdown of protein and polysaccharides by secreted proteases and amylases. Replacement of *Aspergillis* by chemical or enzymatic hydrolysis has no major impact on the organoleptic properties of the sauce. Likewise, inoculation with a pure culture of *Lactobacillus delbrueckii* to carry out the acetic acid fermentation produces a normal product. The *S. rouxii* and *Torulopsis* yeasts, however, are specifically required for proper flavor development.

Microbial Biomass and Single-Cell Protein

In all fermented foods, microbes contribute as preservatives, ie, by lowering the pH and producing ethanol, or by making the food more palatable. The deliberate use of yeasts as food in themselves is less common. Small beer, the sediment from beer, has been traditionally used as a vitamin supplement for infants. Beginning in 1910, dried, spent brewers' yeast was developed as a food, and *Candida utilis* was used as a food supplement in Germany during World War II.

In the 1960s and 1970s, the world protein shortage stimulated the development of additional processes for producing microbial biomass both from traditional substrates, eg, carbohydrates, and from alternative materials such as *n*-paraffins, ethanol, and methanol. Generally, industrial processes for biomass production use yeast cultures, although methanol-based aerobic processes for *Pseudomonas methylotropha* (in the UK) and *Paecilomyces varioti* (in Finland) are also known. Yeasts are preferred both because of their relatively large size compared to bacteria, which permits recovery by centrifugation or filtering more easily, and because of their general acceptance of safety by the food industry. Historically the carbon source for biomass production is fermentable sugars. Because of their low cost and contribution of some nitrogen, minerals and vitamins, beet, and cane molasses are preferred feedstocks. Starch from rice, potatoes, corn, or cassava must be hydrolyzed to produce fermentable sugars. Likewise, cellulose may be used as a carbon source only after acid or enzymatic hydrolysis to glucose. Sulfite waste liquor derived from wood pulp manufacture, lactose-containing whey permeate from cheese-making, ethanol and methanol are also employed. In addition to these carbon sources, yeasts require large supplements of nitrogen, phosphorus, potassium, calcium, and magnesium, as well as trace minerals. Vitamin requirements of different yeasts vary. Bakers' yeast, for example, requires biotin (usually supplied by cane molasses), while *C. utilis* does not (see VITAMINS, BIOTIN).

Like bakers' and brewers' yeasts, yeast for biomass need not be produced under conditions of complete sterility. Most fermentations are carried out at a pH less than 4.5, which limits bacterial growth. The most common contaminants are lactic and acetic acid-producing bacteria. As in the production of bakers' yeast, processes designed for yeast growth are highly aerobic to maximize biomass production and require heat removal. The temperature is typically held at 30°C. The supply of oxygen is also costly, since oxygen transfer rates from air to liquid are low. In many fermentations, the rate of aeration is one volume of air per volume of fermentor liquid per minute. Such a high rate lowers the density of the gas-containing liquid and favors the production of foam, restricting the useful liquid volume of the fermentor.

Continuous processes may be used for the production of yeast biomass. Raw liquid feed is added continuously to the fermentor and an equal volume of fermentor liquid is removed to harvest the yeast cells. These may be a single homogeneous fermentation in a stirred fermentor or two fermentors in series. Growth rates are high: a typical dilution rate in the production of *C. utilis* on sulfite waste liquor is 0.25, ie, one-fourth of the fermentor volume is harvested hourly.

Thus, microbial biomass may be produced either as a by-product of a food fermentation (eg, brewers' yeast) or specifically as a food or feed supplement. In either case, its main importance is in its contribution as a protein supplement. Yeast solids contain about 7.5–9° nitrogen. Protein is crudely calculated as N × 6.25 or about 45–60°, but this includes 6–12° nucleic acids, primarily RNA. A less important contribution consists of the vitamin and mineral content. Also present are 5–9° ash and 2–6° lipids (33). The essential amino acid contents of dry yeasts are quite similar regardless of the species. Lysine content is high, and sulfur-containing amino acids such as methionine are low. The protein efficiency ratio (PER: the ratio of weight gain to weight fed for young rats) of bakers' yeast is 2.02; supplementation with 0.16° D,L-methionine increases this to 2.27, and with 0.5° D,L-methionine to 2.77. Yeast protein purified from most nucleic acids has a PER of 2.2. By comparison, the PER for casein is 2.5.

Brewers' and bakers' dried yeasts are used as dietary supplements. They contribute some protein and trace minerals, and some B vitamins, but no vitamin C, vitamin B₁₂ or fat-soluble vitamins. The glucose tolerance factor (GTF) of yeast, chromium nicotinate, mediates the effect of insulin. It seems to be important for older persons who cannot synthesize GTF from inorganic dietary chromium. The cell wall fraction of bakers' yeast reduces cholesterol levels in rats fed a hypercholesteremic diet.

The most widely available yeast biomass is a by-product of the brewing industry, where the multiplication of yeast during brewing results in a surplus of cells. For every barrel (117 L) of beer brewed, 0.2–0.3 kg of yeast solids may be recovered. In the U.S., a substantial fraction is recovered and made available: about 40 × 10⁶ kg of brewers' yeast annually. The yeast is recovered from beer by centrifuging and dried on roller drums or spray dryers and sold as animal feed or a pet-food supplement. It can be debittered by alkaline extraction to remove the bitter hop residues, and is then sold mainly by the health-food industry. It is available as tablets, powder, or flakes and is often fortified with additional vitamins. Distillers' yeast cannot be readily separated from the fermented mash and the mixture is sold as an animal feed supplement.

Bakers' inactive dry yeast is also widely used in the food industry. This yeast may be grown specifically as a food supplement and consequently there is a choice in its composition by varying growth conditions and feedstock makeup. It can possibly produce high levels of nicotinic acid and thiamin, the crude protein content can be raised to 50–55° and it can be used as a vehicle for the incorporation of micronutrients such as selenium or chromium into the diet.

Candida utilis is grown on sulfite waste liquor in Western Europe and North America, on sugar cane molasses in Cuba and Taiwan and on cellulose acid hydrolysates in Eastern Europe and the former Soviet Union. *C. utilis* utilizes hexoses, pentoses, and many organic acids. Sulfite liquor from hardwoods contains 2–3° fermentable sugars of which 20° are hexoses and 80° pentoses; in softwood liquors the proportions are reversed. The SO₂ must be stripped out to allow yeast growth, which is carried out in large, highly-aerated fermentors. For continuous fermentations, carried out at pH 4 and 30°C, the dilution rate is 0.27–0.30 (34).

Cheese whey solids contain 70–75° lactose, which can serve as the carbon source for lactose fermenting yeasts such as *Khyveromyces fragilis*. The total volume produced is considerably smaller than for the other yeasts described.

Both *n*-alkanes and gas oil can be used as carbon and energy sources. Commercially, *Candida tropicalis* and *Candida lipolytica* have been used (35,36). The fermentation contains two immiscible liquid phases (the alkane and the water); the semisolid yeast; and the gaseous air phase. In contrast to yeasts grown on carbohydrates, where maximum yields are 50°, yeasts grown on alkanes generally give yields of 95–105° based on the weight of the alkane. Cells can be recovered readily by centrifugation but if grown on gas oil must be extracted to remove the residual hydrocarbon. Although some plants operated during the 1970s using alkanes as the substrate, the steep rise in oil prices and regulatory problems have largely stopped this production.

The genera *Hansenula*, *Candida*, *Pichia*, *Torulopsis*, and *Trichoderma* contain facultative methylotrophs. Such yeasts will grow on methanol, carbohydrates or organic acids. Ethanol may be used for biomass production of some species of the genera *Candida*, *Debaromyces*, *Endomycopsis*, *Hansenula*, *Mycoderma*, *Pichia*, *Rhodotroula*, and *Saccharomyces*. In the United States, *C. utilis* is grown on ethanol, and in Finland, bakers' yeast is grown on a combination of ethanol and molasses.

One of the most promising substrates for future production of microbial biomass is the cellulose contained in agricultural residues such as wood pulp, sawdust, feed-lot waste, corn stover, rice hulls, nut shells, and bagasse, all of which contain cellulose as the principal carbon source. Cellulose contents range from 90° in cotton to 15–20° in dicotyledon leaves. Wood residues and grasses contain mixtures of cellulose, hemicellulose, and lignin. The major

impediment preventing commercial use is the difficulty in converting cellulose to glucose by enzymatic or acid treatment. As practiced in Eastern Europe, acid hydrolysis is costly because the acid must be neutralized and the resulting salts disposed of. Enzymatic hydrolysis with cellulase and cellobiase is under intensive research. The most commonly used enzymes are derived from *Trichoderma reesei*. In many instances, 50% of the cellulose can be converted, and higher conversion rates can be achieved by pretreatment of the cellulose with alkali, amines, or ammonia, or by mechanical comminution. Newsprint is an attractive raw material for the production of glucose syrup, ethanol or biomass from the glucose present, but although progress in this area is continuing, these materials are not economical compared with the traditional substrates.

Recovery of biomass from the fermentor requires centrifugation or filtering, washing, concentration, and drying. Yeast grown on sulfite waste liquor contains lignosulfonic acids if not washed. Although suitable for agricultural feed as is, extensive washing is required to produce a yeast suitable for food use.

Extraction of protein requires breaking the cell wall to release the cytoplasmic contents. This can be achieved by high speed ball or colloid mills or by high pressure (50–60 Mpa) extrusion. Protein is extracted by alkaline treatment followed by precipitation after enzymatic hydrolysis of nucleic acids. Although the protein can be spun into fibers or texturized, such products are more expensive than those derived from soybean and there is no market for them.

The presence of nucleic acids in yeast is one of the main problems with their use in human foods. Other animals metabolize uric acid to allantoin, which is excreted in the urine. Purines ingested by humans and some other primates are metabolized to uric acid, which may precipitate out in tissue to cause gout (37). The daily human diet should contain no more than about 2 g of nucleic acid, which limits yeast intake to a maximum of 20 g. Thus, the use of higher concentrations of yeast protein in human food requires removal of the nucleic acids. Unfortunately, yields of protein from extracts treated as described are low, and the cost of the protein may more than double.

Other chemicals of possible concern for health and safety found in yeast proteins include tyramine (0–2.25 mg g) and histamine (0.2–2.8 mg g), formed by decarboxylation of the corresponding amino acids (38). These compounds are also found in other fermented (including pickled) foods. Their presence in yeast extracts used as condiments contributes very little to human intake. Likewise, the nephrotoxic compound lysinoalanine has been identified in alkali-treated yeast extracts, at a level of 0.12 mg g. However, the chemical occurs at similar low concentrations in almost all heat- and alkali-treated foods.

Uses. Inactive dried yeasts are used as ingredients in many formulated foods: baby foods, soups, gravies, and meat extenders; as carriers of spice and smoke flavors; and in baked goods. Yeasts used in the health food industry are generally fortified with minerals and contain higher concentrations of the B vitamins, especially thiamin, riboflavin, and niacin (see VITAMINS).

Dried yeasts are used extensively in animal feeds. Most distillers' and brewers' yeasts are incorporated by co-drying with the spent grains. Approximately 15,000 t of brewers' yeast is dried for this purpose. The main use is in rations for monogastric animals. Active (ie, live) dry yeast is incorporated into feed for ruminants and is believed to aid digestion. Diets supplemented with these yeasts are also claimed to lead to increases in weight gain and productivity. Appreciable amounts of brewers' yeast and torula (*C. utilis*) are also used in pet foods, principally for dogs and cats but also in feeds for birds, fish, mink, and bees.

Yeast-Derived Products

Enzymes. Invertase (β-fructofuranosidase) is commercially produced from *S. cerevisiae* or *S. uvarum*. The enzyme, a glycoprotein, is not excreted but transported to the cell wall. It is, therefore, isolated by subjecting the cells to autolysis followed by filtration and precipitation with either ethanol or isopropanol. The commercial product is available dry or in the form of a solution containing 50% glycerol as a stabilizer. The main uses are in sucrose hydrolysis in high-test molasses and in the production of cream-centered candies.

Lactase (β-galactosidase) is produced commercially from the lactose fermenting *Kluyveromyces fragilis*. The enzyme has a pH optimum of 6–7 and is used in the hydrolysis of lactose in milk or skim milk.

Yeast Extracts. Autolysis in yeast cells is induced by raising the temperature of a cell suspension to 44–55°C; at which temperature, the yeast cells die, but their hydrolytic enzymes remain active. During this process, proteins, carbohydrates, and nucleic acids are hydrolyzed and solubilized. Commercial yeast autolyses generally last over 10 h. At the end of autolysis, the insoluble cell wall materials can be separated from the solubilized products by centrifugation. The resulting extract is evaporated to a paste of about 75% solids or spray dried to a powder. The powder contains 50–80% oligopeptides and amino acids; amino nitrogen may account for 25–45% of the total nitrogen, with the rest consisting mainly of nucleic acids.

Commercially available yeast extracts are made from brewers' yeast, from bakers' yeast, from alcohol-grown yeast (*C. utilis*) and from whey grown yeast (*K. fragilis*). Extracts are used in fermentation media for production of antibiotics, in cheese starter cultures, and in the production of vinegar. They are also extensively used in the food industry as condiments to provide savory flavors for soups, gravies and bouillon cubes, and as flavor intensifiers in cheese products.

The nucleotides inosine-5'-monophosphate and guanosine-5'-monophosphate, produced from yeast RNA are potent flavor potentiators for meat products. They act synergistically with monosodium glutamate and are usually used in conjunction with this amino acid.

Food Preservation and Food Spoilage

Many foods such as alcoholic beverages, pickles, cheese, and fish sauce are preserved by fermentation. Spontaneous fermentations by mixed populations of yeasts and bacteria are normally involved. Preservation results from a lowering of pH or the formation of ethanol. Yeasts do not produce antibiotics, although isolates of a number of species produce a toxin ("killer factor") lethal to other yeasts.

Yeasts also act as spoilage organisms. Jams, jellies, and honey can be fermented by osmophilic yeasts such as *S. mellis* or *S. rouxii*. Wild yeasts may also spoil wines or beers. Film-forming yeasts, such as *P. membranefaciens*, may grow on the surface of sauerkraut, pickles, or other fermented vegetables. *K. fragilis* or other lactose-fermenting yeasts occur in milk products. *C. lipolytica* may occasionally spoil butter or margarine.

Pathogenic Yeasts

Few yeasts are pathogenic to healthy individuals, but immunocompromised persons can suffer from infections from a number of normally innocuous yeasts. *Candida albicans* is the best-known of the pathogenic species. It may cause infections of the skin or of the mucous membranes, eg, in the oral cavity ("thrush"), intestinal tract, and vagina. Occasionally, skin infections may be caused by other species of *Candida*, eg, *C. tropicalis*. *Cryptococcus neoformans* may cause serious infections in a number of organs, particularly the meninges, with sometimes fatal results. Some fungi grow as yeasts at body temperature in laboratory cultures, but grow as mycelia at lower temperatures. Of these, *Histoplasma capsulatum* infects the lungs and may spread to other parts of the body: this systemic disease, histoplasmosis, can be fatal.

Outlook

Historically, yeasts were important in the late nineteenth and early twentieth centuries in elucidating metabolic pathways: the very word "enzyme," introduced by the Buchner brothers, means "in yeast." It is increasingly clear that yeast is of primary importance for the study of eukaryotic cell biology in general: for example, elucidation of the mechanism of cell cycle control and of proteins involved in regulating cell division has been highly dependent on studies undertaken in yeast. The results of these studies could be directly applicable to understanding the mechanism of carcinogenesis in humans. With the complete sequencing of the yeast genome, reported in late 1995, molecular studies will undoubtedly increase. Although genetically engineered yeasts are currently not available for food use, such yeasts have been developed. Commercial fermentations will certainly become more important in the production of feeds, energy, and pharmaceuticals. Genetically engineered yeasts for use in ethanol production are now available (see ETHANOL). Continuous fermentations and fermentations using immobilized cells are under investigation.

BIBLIOGRAPHY

"Yeasts" in *ECT* 1st ed., Vol. 15, pp. 195–219, by W. J. Nickerson, Institute of Microbiology, Rutgers University and A. H. Rose, University of Birmingham, England; "Yeasts" in *ECT* 2nd ed., Vol. 22, pp. 507–554, by George I. deBecze, St. Thomas Institute for Advanced Studies; "Yeasts" in *ECT* 3rd ed., Vol. 24, pp. 771–806, by Gerald Reed, Amber Laboratories, Inc.

1. N. J. W. Kreger-van Rij, ed., *The Yeasts: A Taxonomic Study*, 3rd ed., Elsevier Scientific Publishers, Amsterdam, 1984.
2. J. A. Barnett, R. W. Payne, and D. Yarrow, *Yeasts: Characteristics and Identification*, Cambridge University Press, Cambridge, 1983.
3. A. H. Rose and J. E. Harrison, eds., *The Yeasts*, 3rd ed., Academic Press, New York, 6 vols., 1987–1996.
4. J. Broach, E. Jones, and J. Pringle, eds., *The Molecular and Cellular Biology of the Yeast Saccharomyces*, Cold Spring Harbor Press, New York, 3 vols., 1991–1997.
5. C. L. Cooney in H. J. Rehm and G. Reed, eds., *Biotechnology*, Vol. 1, Verlag Chemie, Weinheim, 1981, p. 73.
6. K. Esser and U. Stahl in H. J. Rehm and G. Reed, eds., *Biotechnology*, Vol. 1, Verlag Chemie, Weinheim, 1981, p. 305.
7. F. Sherman and P. Wakem in C. Guthrie and G. Fink, eds., *Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology*, Vol. 194, Academic Press, New York, 1991, p. 38.
8. S. L. Gerring, C. Connelly, and P. Hieter in C. Guthrie and G. Fink, eds., *Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology*, Vol. 194, Academic Press, New York, 1991, p. 57.
9. C. Guthrie and G. Fink, eds., *Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology*, Vol. 194, Academic Press, New York, 1991.
10. K. K. Kapoor, K. Chaudhury, and P. Tauro in G. Reed, ed., *Prescott and Dunn's Industrial Microbiology*, 4th ed., AVI Publishing Co., Westport, Conn., 1982.
11. K. Nakayama in Ref. 10.
12. Y. Kyogoku and K. Ouchi, *Appl. Env. Microbiol.* **61**, 639 (1995).
13. P. L. Altman and D. S. Ditmar in *Metabolism, Handbook of FASEB*, FASEB, Bethesda, Md., 1968.
14. J. White, *Yeast Technology*, Chapman and Hall, London, 1954.
15. "Basic Yeast," Universal Foods, Milwaukee, 1992.
16. G. Reed and T. W. Nagodawithana, *Yeast Technology*, 2nd ed., Van Nostrand Reinhold, New York, 1991.
17. M. J. Wolt and B. L. D'Appolonia, *Cereal Chem.* **61**, 209 (1984).
18. E. J. Pyler, *Baking Technology*, 3rd ed., Sosland Pub. Co., Merriam, Kans., 1988.
19. T. F. Sugihara, L. Kline, and M. W. Miller, *Appl. Microbiol.* **21**, 456 (1971).
20. L. Kline and T. F. Sugihara, *Appl. Microbiol.* **21**, 459 (1971).
21. D. Samuel, *Nature* **273**, 488 (1996).
22. S. R. Griffin, *J. Inst. Brewing* **76**, 45 (1970).
23. M. Amaha, *Bull. Brewing Sci. (Japan)* **12**, 43 (1966).
24. I. C. MacWilliam and J. F. Chatterton, *Eur. Brew. Conv. Proc.*, 271 (1969).
25. P. E. McGovern, D. L. Glusker, L. J. Exner, and M. M. Voight, *Nature* **381**, 480 (1996).
26. R. T. DeSoto, *Am. J. Enol. Vitic.* **6**, 26 (1955).
27. B. C. Rankine, *Ann. Technol. Agric.* **27**, 189 (1978).
28. B. C. Rankine, *J. Appl. Sci.* **4**, 590 (1953).
29. S. C. Pan, A. A. Andreasen, and P. Kolachov, *Ind. Eng. Chem.* **42**, 1783 (1950).
30. K. Kodama in A. H. Rose and J. E. Harrison, eds., *The Yeasts*, Vol. 3, Academic Press, New York, 1970, p. 225.
31. T. Inoue, Y. Takoaka, and S. Hata, *J. Ferment. Technol. (Osaka)* **40**, 544 (1970).
32. F. M. Yong and B. J. Wood in D. Perlman, ed., *Adv. Microbial Microbiol.*, Academic Press, New York, **17**, 157 (1974).
33. R. Kihlberg, *Annu. Rev. Microbiol.* **26**, 427 (1972).
34. H. J. Peppler in Ref. 30, p. 421.
35. R. G. Cooper, R. S. Silver, and J. P. Boyle in S. R. Tannenbaum and D. I. C. Wang, eds., *Single Cell Protein*, Vol. 2, MIT Press, Cambridge, Mass., 1975.
36. K. Allison in F. Wagner, ed., *Microbial Production of Protein*, Chemie Verlag, Weinheim, 1975, p. 73.
37. C. I. Waslien and co-workers, *J. Food Sci.* **228**, 180 (1970).
38. J. M. Concon, *Food Toxicology, Principals and Concepts*, Marcel Dekker, New York, 1988, pp. 622–627.

Daniel Maloney
Universal Foods Technical Center

YOUNG-HELMHOLTZ COLOR-VISION THEORY.

See COLOR.

YTTERBIUM.

See LANTHANIDES.

ZEOLITES.

See MOLECULAR SIEVES.

ZIEGLER-NATTA CATALYSTS.

See CATALYSIS; OLEFIN POLYMERS.

ZINC AND ZINC ALLOYS

Zinc [7440-66-6] is a relatively active metal and its compounds are stable. Since it is not found free in nature, it was discovered much later than less-reactive metals such as copper, gold, silver, iron, and lead. In early times, the smelting of copper ores containing zinc resulted in brasses, which were known to the Romans before 200 BC. Later, brasses were made by heating copper with zinc oxide or carbonate and charcoal. The oldest piece of zinc extent is an idol found in a prehistoric Dacian site in Transylvania which analyzed at 87.52° zinc, 11.41° lead, and 1.07° iron (1). Zinc smelting is thought to have originated in China, where it was known in the seventh century AD how to make malleable zinc. In India, zinc was produced from ore mined at Zawar before 1380. By the seventeenth century, zinc was imported into Europe from Asia and in 1743 a zinc smelter for zinc oxide ore was erected in Bristol, UK. By the early nineteenth century, zinc smelting was well established in Germany and Belgium. Zinc was first produced in the United States at the arsenal in Washington, D.C., in 1835, and by 1860 The New Jersey Zinc Company had well-established smelting operations at Bethlehem, Pennsylvania.

In 1758 in England, John Champion was granted a patent for making zinc from roasted zinc sulfide, the main ore of zinc. This route is still the principal method of production today. Commercial zinc metallurgy started with batchwise horizontal retorts, which have had many variations, and progressed to continuous vertical-retort processes. Although some of the latter are important today, electrowinning dominates the field, and the Imperial Smelting blast furnace for zinc–lead remains competitive.

The main application of zinc is to protect iron and other metals from corrosion. Zinc in contact with iron and other metals, as a coating or attached anode, corrodes sacrificially and protects the iron. Most commonly, zinc is applied in the molten state, ie, galvanizing, but is also applied either by electrodeposition or by various mechanical procedures using zinc dust or powder (see METALLIC COATINGS; METAL SURFACE TREATMENTS). Zinc dust paints are growing in importance (2). Another important use is in alloys for die casting. These alloys are used extensively because of their high quality and low cost. Brass and bronze products account for the third largest usage.

The monograph on zinc is a valuable general reference on zinc technology (3). Furthermore, detailed descriptions of extractive processes, resource data, and environmental- and energy-related papers from symposia of the Metallurgical Society of the AIME are a rich source of information (4–7).

Occurrence

Zinc, like most metals, is found in all natural waters and soils as well as the atmosphere and is an important trace element in plant and animal life (see MINERAL NUTRIENTS). Rocks of various kinds contain 20–200 ppm zinc and normal soils 10–30 ppm (average ca 50 ppm) in uncontaminated areas. The average zinc content of coal is 33 ppm. Seawater contains 1–27 µg L (median ca 8 µg L), and uncontaminated freshwater usually <10 µg L.

Zinc ores are widely distributed throughout the world; 55 zinc minerals are known (8–10). However, only those listed in Table 1 are of commercial importance. Of these, sphalerite provides ca 90° of the zinc produced today. Sulfide ores usually occur in the range of 2–12° zinc (average ca 4°) as mined.

Table 1. Common Zinc Minerals

Name	Composition	° Zn
sphalerite ^a	ZnS	67.0
hemimorphite ^b	Zn ₄ Si ₂ O ₇ (OH) ₂ ·H ₂ O	54.2
smithsonite	ZnCO ₃	52.0
hydrozincite	Zn ₅ (OH) (CO ₃) ₂	56.0
zincite	ZnO	80.3
willemite	Zn ₂ SiO ₄	58.5
franklinite	(Zn,Fe,Mn)(Fe,Mn) ₂ O ₄	15–20

^a Zinc blend, wurtzite.

^b Calamine.

The oxygen-containing ores in Table 1 are often found in richer deposits ranging up to 35% zinc, most commonly smithsonite and hemimorphite. They are thought to have originated from sulfide mineralization and are minor sources of zinc. However, in the beginning, the U.S. zinc industry was based on the deposit of franklinite, zincite, and willemite in Sussex County, New Jersey. Zinc deposits are classified as contact metamorphic, irregular and associated fissure fillings, vein, stratabound in metamorphic rocks, stratabound in carbonate rocks, strataform, and deposits formed by supergene enrichment of laterization (10). Carbonate rocks, eg, limestone and dolomite, are common host-rocks of zinc ore.

In the United States, the richest zinc district is in Alaska; however, Australia leads the world by every measure of reserves followed by Canada and the United States (see Tables 2 and 3).

Table 2. Western World Zinc Reserves, Contained Zinc in 1996, Existing Mines and Advanced Projects^a

Country	Proven and probable	Indicated and inferred	Total
Australia	32.4	20.2	52.6
Canada	21.0	3.8	24.9
United States	17.6	13.8	31.4
Peru	14.2	3.1	17.4
India	9.5	1.4	10.9
Iran	5.8	0.0	5.8
Mexico	5.1	2.4	7.5
Spain	5.1	1.5	6.6
Eire	4.4	1.3	5.8
Zaire	3.1	2.6	5.6
Namibia	1.8	0.0	1.8
Sweden	1.8	0.8	2.6
Brazil	1.4	0.0	1.4
Morocco	1.1	1.2	2.3
Bolivia	1.0	0.0	1.0
Greece	0.7	0.8	1.6
Turkey	0.7	0.5	1.2
Saudi Arabia	0.7	0.1	0.8
South Africa	0.6	0.0	0.6
Japan	0.4	0.0	0.4
Tunisia	0.4	0.0	0.4
France	0.3	0.0	0.3
Honduras	0.3	0.0	0.3
Argentina	0.3	0.0	0.3
Thailand	0.2	0.0	0.2
Chile	0.2	0.0	0.2
United Kingdom	0.2	0.2	0.3
Finland	0.1	0.0	0.2
Norway	0.04	0.0	0.1
Italy	0.03	0.0	0.0
Total	130.47	54.03	184.51

^a Source: Brook Hunt & Associates, Addlestone, Surrey, U.K.

Table 3. U.S. Zinc Reserves^a

District	States	Reserves
Alaska	Alaska	15.7
Mississippi Valley	Illinois, Missouri, Wisconsin, mid-Tennessee	5.8
South Appalachian	Virginia, east Tennessee	6.5
Northeastern	Maine, New York, New Jersey, Pennsylvania	3.4
Southwestern	Arizona, New Mexico, Utah, Colorado, California	2.1
Northwestern	Idaho, Montana, Washington	1.2
Total		34.7

^a Ref. 11.

Zinc minerals tend to be associated with those of other metals; the most common are zinc–lead or lead–zinc, depending upon the dominant metal, zinc– copper or copper–zinc, and base metal such as silver. Zinc does occur alone, most often in the northeastern district, and here, as elsewhere, recoverable amounts of cadmium (up to 0.5%) are present. Other minor metals recovered from zinc ores are indium, germanium, and thallium.

The exploration, evaluation, and development of zinc and lead ore bodies in North and Central America are discussed in Ref. 12. A survey of world zinc production in Ref. 13 gives all operating mines and mills, and their methods, production, and chemical analysis of the products; zinc smelters are included.

Open-pit zinc mining is not common, since most mines are below the surface. The Kidd Creek Mine in Ontario, Canada, is a combination open-pit–underground mine. It is one of the richest deposits in the world with an estimated 62.5×10^6 t grading 7.08% zinc, 1.33% copper, and 151 g silver (14). Underground mining methods include room-and-pillar, shrinkage, cut-and-fill, and square set. In the United States, ca 20 mines account for more than 98% of zinc production.

Physical Properties

Zinc is a lustrous, blue-white metal, which can be formed into virtually any shape by the common metal-forming techniques such as rolling, drawing, extruding, etc. The hexagonal close-packed crystal structure governs the behavior of zinc during fabrication. Physical properties are given in Table 4.

Table 4. Physical Properties of Zinc^a

Property	Value
----------	-------

ionic radius, Zn ²⁺ , nm	0.074
covalent radius, nm	0.131
metallic radius, nm	0.138
ionization potential, eV	
first	9.39
second	17.87
third	40.0
density	
solid, g cm ³	
at 25°C	7.133
at 419.5°C	6.830
liquid, g mL	
at 419.5°C	6.620
at 800°C	6.250
melting point, °C	419.5
boiling point, °C	907
heat of fusion at 419.5°C, kJ mol ^b	7.387
heat of vaporization at 907°C, kJ mol ^b	114.8
coefficient of thermal expansion, mm (m·K)	
volume	8.9
linear, polycrystalline	39.7
thermal conductivity, W (m·K)	
solid	
at 18°C	113.0
at 419.5°C	96.0
liquid	
at 419.5°C	60.7
at 750°C	56.5
electrical resistivity, nΩ m	
polycrystalline	$R = 54.6 (1 + 0.0042 t)$ ^c
liquid at 423°C	369.55
heat capacity, J (mol·K) ^b	
solid	$22.39 + 10^{-2} T$ ^d
liquid	31.39
gas	20.80

^a Ref. 15.

^b To convert J to cal, divide by 4.184.

^c $t = 0\text{--}100^{\circ}\text{C}$

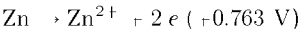
^d $T = 298\text{--}692.7\text{ K}$.

Mechanical history, heat, and impurities greatly affect the mechanical properties. Pure zinc is ductile at room temperature and does not have a definite yield point as do most structural metals. Rather, it creeps under sufficient constant load. The impurities of commercial zinc and alloying metals are carefully controlled to achieve the desired mechanical properties.

Chemical Properties

The most significant chemical property of zinc is its high reduction potential. Zinc, which is above iron in the electromotive series, displaces iron ions from solution and prevents dissolution of the iron. For this reason, zinc is used extensively in coating steel, eg, by galvanizing and in zinc dust paints, and as a sacrificial anode in protecting pipelines, ship hulls, etc.

In batteries, a zinc anode undergoes the oxidation reaction,



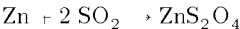
to provide a flow of electrons to the external circuit. For instance, in the familiar dry cell, the zinc is oxidized to zinc chloride while manganese dioxide is reduced at the cathode. The alkaline zinc battery with manganese dioxide cathode uses potassium hydroxide electrolyte and forms zincate ions or precipitates zinc oxide at the anode:



Other alkaline primary cells couple zinc with oxides of mercury or silver and some even use atmospheric oxygen (zinc–air cell). Frequently, zinc powder is used in the fabrication of batteries because of its high surface area. Secondary (rechargeable) cells with zinc anodes under development are the alkaline zinc–nickel oxide and zinc–chlorine (see BATTERIES).

The capability of zinc to reduce the ions of many metals to their metallic state is the basis of important applications. However, metals are removed from zinc solutions by displacement with finely divided zinc before winning by electrolysis. Gold and silver are displaced from cyanide leach solutions with zinc and the following metals are similarly recovered from various solutions: platinum group, cadmium, indium, thallium, and sometimes copper.

Zinc hydrosulfite (zinc dithionite) is a powerful reducing agent used in bleaching paper and textiles; it is prepared from zinc dust and sulfur dioxide:



Another hydrosulfite reducing agent is zinc formaldehyde sulfoxylate, Zn(HSO₂·CH₂O)₂ (see BLEACHING AGENTS).

Pure zinc displaces hydrogen from acidic solutions slowly because of its high hydrogen overvoltage. This fact allows zinc to be electrodeposited by reduction from highly purified acidic solutions with only slight evolution of hydrogen. Surface amalgamation with mercury is useful in alkaline-battery applications because it increases the hydrogen overvoltage and reduces hydrogen gassing. Conversely, some metals reduce overvoltage and increase the rate of hydrogen evolution. Zinc dust, because of its high surface area, must be kept bone dry to avoid slow displacement of hydrogen from water. Massive zinc, on the other hand, reacts with steam rapidly above 350°C (see also ELECTROCHEMICAL TECHNOLOGY; ELECTROPLATING).

Oxidizing elements such as oxygen, sulfur, and halides react with zinc at room temperature in the presence of moisture, but do not in its absence. At higher temperature, the reactions can be vigorous even when dry. For instance, a powdered mixture of zinc and sulfur explodes if warmed and zinc reacts

rapidly with oxygen at 225°C. Atmospheric corrosion of zinc results in hydrated basic carbonates, $x\text{ZnO} \cdot y\text{CO}_2 \cdot z\text{H}_2\text{O}$, of variable stoichiometry which form a surface film.

Zinc does not react with nitrogen, even at elevated temperatures but zinc nitride, Zn_3N_2 , forms with ammonia at red heat. Zinc sulfide, the most common form of zinc in nature, is not reduced directly in commercial practice because of reactions of the zinc vapor during condensation. Rather, the sulfide is burned (roasted) to the oxide plus sulfur dioxide before reduction. However, zinc can be reduced to the metal at ca 1300°C with carbon or iron.

In the Parkes desilvering process, 1–2% zinc is added to molten lead where it reacts with any gold, silver, and copper to form intermetallic compounds which float as crusts or dross that is skimmed (see LEAD AND LEAD ALLOYS).

Processing

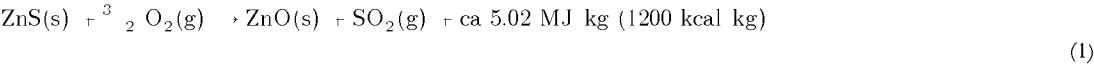
Concentration. Zinc ores are too low in zinc content for direct reduction and must be concentrated. They are first crushed, usually underground, in jaw, gyratory, or cone crushers, and then ground to 75–150 μm (100–200 mesh) by ball or rod milling to a degree necessary to separate the zinc minerals from other minerals and gangue in later flotation (qv). In some ores, the zinc and lead minerals are too intimately mixed to be separated by flotation and are separated by leaching for subsequent electrowinning or in the Imperial Smelting furnace (ISF). The beneficiation of zinc ores is usually more complex than for other nonferrous ores because of the diversity of minerals present. Not only are copper or lead often associated, but the sulfide ore is sometimes partially oxidized.

Beneficiation before grinding can be done by gravity or magnetic methods (16). Wetherill high intensity magnetic separators were developed by The New Jersey Zinc Company in 1896 to separate franklinite from their Franklin and Sterling ores (see MAGNETIC SEPARATION). Jigging and tabling recovered the zincite and willemite. The concentrates were further upgraded to crude zinc oxide (17).

Zinc ores are generally floated at the mine (18). In the case of simple zinc sulfide ores, flotation is carried out by treatment with copper sulfate to activate the sphalerite causing it to be wet by the organic collector (eg, xanthate). The now-hydrophobic zinc ore particles attach themselves to the rising bubbles. Oxidized ore particles present must be sulfidized with sodium sulfide to be floated (19). Flotation produces concentrates which are ca 50–60% zinc. In mixed ore, the lead and copper are usually floated after depressing the sphalerite with cyanide or zinc sulfate. The sphalerite is then activated and floated.

Roasting. Copper and lead sulfides are directly smelted but not zinc sulfide. However, theoretical calculations are encouraging (20) and, if an efficient means of condensing zinc rapidly from 1600 K in the presence of carbon dioxide, sulfur dioxide, and steam can be devised, the process may be feasible. The reaction of zinc vapor to yield zinc oxide or zinc sulfide presents the main difficulty.

The zinc sulfide in the concentrate is always converted to oxide by roasting. An exception is the direct leach process described below. The principal overall roasting reaction is strongly exothermic and provides excess heat which is recovered.



Under certain conditions, the oxide is sulfatized:



In a usual roaster gas at equilibrium, the sulfate decomposes at ca 860°C, but it is difficult to avoid some sulfatization during cooling and indeed some plants require a degree of sulfation to maintain sulfate balance.

Another important reaction accounts for sulfur trioxide in the roaster off-gases and is a component of reaction 2.



Reaction 3 also occurs on cooling since the concentration of SO_3 is very low at roaster temperatures of 950°C and approaches zero at 1000°C. Another important reaction that occurs during roasting is the formation of zinc ferrite, $\text{ZnO} \cdot \text{Fe}_2\text{O}_3$, above 650°C (see FERRITES). Zinc ores contain 5–12% iron. Zinc ferrite forms solid solutions with other spinels, such as $\text{FeO} \cdot \text{Fe}_2\text{O}_3$, and therefore the zinc–iron compositions formed are of indefinite stoichiometry. Ferritic zinc is difficult to solubilize in hydrometallurgical leaching but several recovery processes are discussed below.

The sulfur dioxide of reaction 1 is cooled in a waste-heat boiler, freed from calcine, and converted to trioxide. The oxidation and conversion to sulfuric acid is conducted in a conventional acid plant (see also SULFURIC ACID AND SULFUR TRIOXIDE).

For environmental and economic reasons, the early practice of roasting zinc sulfide and discharging the sulfur dioxide to the atmosphere gave way to plants where the sulfur dioxide is converted to sulfuric acid. Desulfurization takes place while the ore particles are suspended in hot gases. Called flash- and fluid-bed roasters, these processes are described below. Some plants use combinations of roasters and sintering for desulfurization.

Multiple-Hearth Roasters. The circular types consist of a series of hearths arranged vertically in such a way that the ore entering the top is rabbled and dropped down from hearth to hearth, until it is completely oxidized. The hearths are usually stationary and the plows revolve, such as in the Wedge, Herreshoff, Ord, Skinner, and other roasters (21). In other furnaces, the hearths revolve and the rabbles are fixed, eg, the deSpidlet and its modification, the Barrier.

A conventional circular-wedge roaster consists of a brick-lined steel shell with hearths arched gently upward from the periphery to a central shaft. The brick hearths may number from 8 to 16 and are ca 1 m apart. The central steel shaft (ca 1.2 m in diameter) revolves at 1 rpm or less carrying two rabble arms per hearth. These rabbles, cooled with air or water, plow the ore from the outside to the center of the hearth where it is dropped to the next hearth for plowing in the opposite direction. The calcine thus proceeds to the bottom where it is dropped into a conveyor. The sulfide sulfur at this point is ca 3.5% (22).

The uppermost hearth serves to dry the damp ore in the hot (ca 500°C) gases exiting the top of the toaster. These gases may contain up to 15% of the total crude oxide and up to 6% sulfur dioxide, high enough to be fed to a sulfuric acid plant. In some pyrometallurgical operations, desulfurization is continued in a sintering step.

Multiple-hearth roasting offers ease of operation, ability to handle a wide variety of ores or blends, and little downtime. On the other hand, these furnaces are no longer being built because of their high capital and labor costs, relatively low sulfur dioxide off-gas, need for added fuel, and marginal opportunity for waste-heat recovery.

Flash Roasting. As more and more zinc concentrates were produced by flotation, which requires finely ground ore, more rapid roasting was possible. Instantaneous oxidation was noted in multiple-hearth furnaces when particles were dropped from one hearth to another. The Consolidated Mining and Smelting Company of Canada (Cominco, Ltd.) took advantage of this observation to develop the flash or suspension roaster at Trail, B.C., Canada, in the early 1930s (23). An 8-m Wedge roaster with eight hearths was modified by removing all but the top two and bottom two hearths. Ore, fed to the top, is dried and then removed for grinding. After classification, it is conveyed back to the roaster where it is blown into the central chamber in a turbulent fashion using high pressure air jets. Oxidation occurs at ca 1000°C and ca 40% of the product falls to the bottom hearths. This coarser material is further desulfurized as it is rabbled out of the roaster. Some or all of the dust leaving with the gas stream is usually returned to the bottom hearths for further oxidation or decomposition of sulfate. The dust is removed from the gas by means of a cyclone following the waste-heat boiler. An electrostatic precipitator, wet scrubber, and dryer (sulfuric acid tower) clean up the gas before it passes to the acid plant.

New flash roasters dry on the bottom hearth; the ore is introduced in two opposed burners for increased turbulence (24). Such roasters with combustion chambers of 8–9 m high are capable of dead roasting (sulfide removal to <0.5%) over 300 t of zinc concentrates per day with 10% sulfur dioxide in the off-gas.

The flash roaster is flexible in handling various flotation concentrates and reaching the degree of desulfurization desired, ie, 0.5–3.0% sulfate sulfur. Waste heat is easily recovered. However, grinding and rabbling must be done mechanically.

Fluidization Roasting: Fluid-bed roasting offers the advantages of reduced machinery, high throughput, good heat recovery, and rich sulfur dioxide off-gas (25,26) (see FLUIDIZATION). Most electrolytic zinc plants use fluid-bed roasting because of the ease of controlling the temperature to achieve proper sulfide and sulfate levels. Compared to flash roasting, zinc leachability is better and sulfation is increased while the formation of zinc ferrite is minimized (see also EXTRACTIVE METALLURGY).

Slurry Fed. Dorr-Oliver, Inc., pioneered fluidization roasting and developed their FluoSolids process originally for treating arsenopyrite gold ores. The reactor (27) for roasting up to 155 t of concentrates per day is an insulated cylinder of brink-lined steel construction and with a conical bottom 9.5 m high with a diameter of 6.3–7.5 m. The windbox is separated from the reaction chamber by a distribution plate containing 780 nozzles designed to keep calcine from falling into the windbox. Fluidization air is supplied at 20,000 m³ h at 34.3 kPa (257 mm Hg), ie, 30% excess over stoichiometric. In operation, the bed is 1.5 m deep and at 1.5 kPa (11 mm Hg), under a slight positive pressure.

After start-up, the temperature is raised to the bed ignition temperature of 900°C. The feed slurry at 80% solids is pumped from an agitated tank and injected by atomization with compressed air. Evaporation of water from the slurry, or added supplementally, controls roasting at 900°C. The coarse calcine (40% of total) exits the top of the bed and flows to an evaporative cooler followed by cyclone and bag filter. The gases carry 60% of the solids to a waste-heat boiler where 15% of the total calcine is collected. Cyclones recover 40%, and 5% is picked up by an electrostatic precipitator. The gas at this point is ca 10% sulfur dioxide. The collected calcine is screened and the coarse fraction (>2 mm) is hammer-milled before being sent to leaching in the electrolytic plant.

The coarse calcine cooler operates at 300°C, while the waste-heat boiler cools the gas to 350°C. The tubes in the boilers have a chain-shaking arrangement operated by pneumatic hammers. Steam production is 0.78 kg/kg of dry concentrate. The only trouble with dust reported is in the connection between the reactor and waste-heat boiler. It is necessary to cool the gas stream quickly to avoid sulfation, but even so the carry-over calcine contains on the order of four times more sulfate than the coarse overflow. In this plant, the composite calcine is 0.1% sulfide and 2.2% sulfate sulfur.

A similar FluoSolids roaster handles 155 t/d of zinc concentrates through a roaster which is 1 m higher than standard (28).

The Dorr-Oliver slurry-fed roaster allows easy introduction of the concentrate, relatively low dust carry-over, good sealing, and freedom from cooling coils in the bed. It is usually operated at negative pressures to avoid sulfur dioxide and dust leakage but the in-leakage of air causes sulfation which may be undesirable. A temperature >900°C aids desulfurization by reducing carry-over; that is, the higher temperature induces particle sintering and agglomeration which diverts more of the calcine to bed overflow where it avoids sulfation. The principal disadvantage of the slurry-feed system is that the added water reduces the amount of heat that can be recovered.

Dry Fed. The dry feeding of fine concentrate to fluid-bed roasting was developed at S.A. Vieille-Montagne (VM) in Balen, Belgium, using a modified turbulent-layer process (BASF pyrites roaster). This process was sold to the zinc industry by Lurgi whose largest installation treats 800 metric tons of concentrates per day (29). In this roaster, unlike the Dorr-Oliver FluoSolids, so-called dry concentrates with 5–10% moisture are fed by mechanical belt slingers. In addition, the freeboard diameter is larger. Heat is removed from the bed by cooling coils and adding water. The VM roaster is the best for heat recovery and has a relatively low gas throughput.

The newest zinc plant in the United States at Clarksville, Tennessee, owned by Savage Zinc Company, uses a VM roaster (30). A single roaster treats 382 t of concentrate per day and recycles dross producing 333 t of calcine, while the acid plant produces 338 t of 100% sulfuric acid. The grate area of the roaster is 64.7 m² and normally would have five cooling coils in the bed. The low heat value (low iron sulfide) of the concentrates allows removal of three coils. Calcine leaves the roaster as bed overflow (ca 60%) or carry-over into the waste-heat boiler (ca 40%). The latter generates 18,160 kg of steam per hour at 3.9 MPa (ca 39 atm) pressure. Gases leaving the electrostatic precipitator contain 9.5% SO₂. Calcine from the roaster overflow and waste-heat boiler are cooled to ca 100°C in a rotary-drum cooler and combined with cyclone and precipitator collects for dry ball-milling to <120 μm. Sulfur levels in the calcine are 0.25% sulfide and <1.5% sulfate. The clean, cool gas passes through a Norzinc mercury-removal tower (31) before entering the acid plant. Mercury reacts with soluble mercury(II) chloride to precipitate mercury(I) chloride, ensuring <1 ppm mercury content.

Canadian Electrolytic Zinc, Ltd., at Valleyfield, Quebec, uses two VM roasters to process 360 t of concentrates per day (32). Each roaster has 3,300 tuyeres in its grate of 34 m² area, to fluidize a bed of 1 m depth; gas exits at 990°C. This plant was the first to use a water-cooled settling chamber in the waste-heat boiler for reducing the degree of sulfation. The gas exits the boiler at 320°C. An attempt was made to use spent electrolyte from the cells for roaster temperature control. This was designed to control the total sulfate in the plant but it caused excessive bed agglomeration and increased sulfation.

Agglomerate Fed. Fluid-bed roasters of this type were developed in order to avoid the sintering stage in preparing feed for pyrometallurgical zinc operations where agglomerate strength and freedom from volatile components (sulfates, cadmium, etc) is essential.

Metallurgie Hoboken-Overpelt S.A. (MHO) in Belgium uses hard pellets as a suitable feed for horizontal retorts (33). The MHO roaster, started in 1954, was the first to recover waste heat and was developed in collaboration with Dorr-Oliver. Concentrates and recycled fines are homogenized with sulfuric acid from scrubbing and pelletized to 0.5–4-mm size. The pellets are dried and fed to the rectangular roaster where the bed flow is longitudinal to the discharge port. The furnace is 5.5 m high inside. For 85% of the length, fluidization air is introduced and the balance of the bottom is an aperture where the pellets fall into a finishing bin. Crude pellets are less dense than calcine and tend to float in the upper portion of the 1-m-deep bed. Calcine enters the finishing bin at 980°C where incoming air completes the oxidation and sweeps away the sulfur dioxide to prevent sulfation. Sulfur content of the calcine is ca 0.05% sulfide and 0.35% sulfate. The dust problem is reduced by carrying over only 20% of the calcine to the boiler. A relatively long retention and high temperature (up to 1060°C) tend to eliminate chlorides, fluorides, mercury, and selenium. Ores containing low melting materials such as silica and magnesia do not create sticking problems. A disadvantage is the need for considerable grinding before leaching.

The New Jersey Zinc Company patented a fluidized-pellet roaster which was installed in several zinc plants. Called a fluid-column roaster, it resembles a shaft furnace and can handle 370 t of concentrate per day. This roaster can be operated at 1080–1100°C to eliminate 90% of the cadmium and 92% of the lead. The fluid-column roaster has the same advantages as the MHO roaster; the pelletizing cost is a disadvantage for both systems.

St. Joe Minerals Corporation uses a fluid-bed roaster to finish the roasting at 950°C of material that has been dealed in a modified multiple-hearth furnace operated with insufficient oxidation (34). First, sulfur is reduced from 31 to 22% and lead from 0.5 to 0.013%. Somewhat aggregated, the product is hammer-milled before final roasting. Half of the calcined product is bed overflow and special hot cyclones before the boiler remove the other half; total sulfur is ca 1.5%. Boiler and precipitator dusts are higher in sulfur, lead, etc, and are separated.

Sintering: Sintering before further pyrometallurgical smelting completes the roast, eliminates volatile material, and aggregates fine calcine. Aggregates need considerable strength and porosity in vertical retorts, electrothermic furnaces, and the Imperial Smelting process (ISF). Sintering practice can be classified as follows (22): on low sulfur calcine with carbonaceous fuel in one or two stages; on unroasted concentrates with high sinter recycle; and on 7–9% sulfur calcine without fuel. Zinciferous material is mixed with carbonaceous fuel, if necessary, and water. The mix is pelletized, or partly so, and fed onto a traveling endless grate. After ignition, air is drawn through the bed to burn the carbon and zinc sulfide.

Sintering machines, usually of the Dwight-Lloyd downdraft type, are used. In downdraft operations, metallic impurities are usually not eliminated in one pass because they tend to condense in the bottom portion of the bed. For this reason, the product is sliced from the top of the 10–20-cm bed and the bottom fraction (20–40%) recycled or used where the impurity can be tolerated. Chloride may be added to improve the transfer of cadmium and lead from the bed into the windbox and downstream dust and fume-collection system. Since the electrothermic zinc process requires very strong pellets, double sintering is practiced with silica flux added to the second stage. Heat is not recovered from downdraft sintering operations because of high gas volumes, moisture in the charge, and large sinter recycle loads, and a considerable energy debit is incurred. In addition, it is expensive to scrub the dilute sulfur dioxide in the off-gases.

Updraft sintering is used on unroasted concentrates and the off-gases are built to a high enough sulfur dioxide content (5–6.5%) to allow sulfuric acid production. This is accomplished by preventing gas leakage around the bed with grease-sealed slide rails and by gas recirculation. In order to avoid excessive bed temperatures and maintain porosity, sinter recycle (ca 80%) is high. Some heat is recovered to heat incoming air.

Reduction.

Electrolytic Process. The electrolytic process, where zinc is deposited from an aqueous solution onto the cathode, treats complex ores that did not lend themselves to pyrometallurgical recovery (see also ELECTROCHEMICAL PROCESSING). For instance, some abundant ores containing combinations of lead, copper, silver, iron, and zinc could be concentrated to only 30–40% zinc, an unsuitable level for pyrometallurgy. The first successful electrolytic zinc plant was built by the Anaconda Company at Anaconda, Montana, and began production in 1915. The sulfide concentrate is oxidized to crude oxide and leached with return acid from the cells. The zinc sulfate solution is purified and electrolyzed. Some plants also feed crude zinc oxide from pyrometallurgical recovery operations. At present, there are four electrolytic zinc plants in the United States with annual capacities from 50,000 to 100,000 metric tons. A simplified flow diagram is shown in Figure 1 and detailed descriptions of plants and processes are given in the literature (35–39).

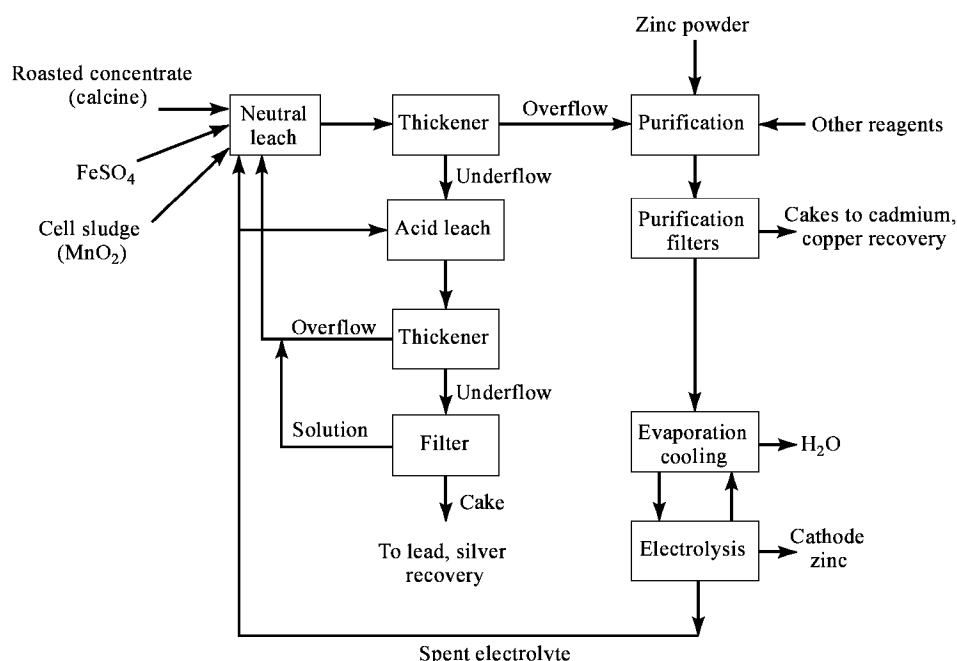


Fig. 1. Simplified flow diagram, electrolytic zinc.

Leaching. Crude zinc oxide in the form of roasted concentrate and, at some plants, fume from slag-fuming is fed to the leach circuit where zinc is solubilized with 100–200 g/L sulfuric acid which has been returned from the cells. Grinding is integrated into the leaching step. Zinc ferrite formed in roasting is not soluble in the dilute acid and this zinc is recovered. Residues also contain lead and often silver. The pH of the solution from which the residue is separated is ca 5.0 where hydrous iron(III), aluminum, indium, and silicon oxides are precipitated but basic zinc sulfate is not. The hydrous oxides adsorb arsenic, antimony, germanium, and other impurities. Thus, leaching is important in purification.

Leaching is a one- or two-step procedure, batch or continuous. In one-step leaches, calcine is added to neutralize return acid until the soluble zinc is in solution and the pH is high enough to precipitate iron(III) and other impurity hydroxides. The final pH adjustment may be made with lime or limestone instead of calcine to decrease zinc loss to the residue. Although the one-step procedure is simple and requires less capital equipment than the two-step, it must be carefully controlled to ensure zinc dissolution and avoid precipitation of basic zinc sulfate above pH 5.5. Residues containing 15–20% zinc (primarily ferrite) are sent for pyrometallurgical recovery. In the two-step procedure, more impurities are removed. New zinc plants use various modifications of this procedure because it includes recovery of the ferritic zinc. The first leach of the two-step procedure is called the neutral leach. Excess calcine is added to ca half the return acid plus acidic solution from the second leach step to precipitate iron and other impurities and ensure a pH (5.0–5.2) high enough for proper flocculation of the residue. The leachate is sent to the purification section and the residue, containing ca 50–75% of the available zinc, is leached with acid to dissolve the remaining available zinc. The final acid contains 3–5 g zinc per liter and, although some of the impurities remain in the solution returning to neutral leach, most remains with the residue. Depending on the ore, the residue may also contain lead, copper, silver, and gold. In many plants, the hot acid leach (95°C) dissolves zinc ferrites and other values which otherwise would result in the loss of up to 15% of the zinc and higher proportions of cadmium and copper.

The leaching is 50–60°C without external heating. The cone-bottom tanks are equipped with a pipe from just above the solution level to near the bottom through which air is blown forming bubbles which lower the density of the slurry. These leach tanks, called Pachucas, are fairly efficient and are still popular although many plants employ mechanical agitators. Most plants use 3–5 tanks in series with acid and calcine being fed to the first and, in some cases, downstream from the first tank as well.

Recovery of Zinc from Leach Residue. The residues from dilute acid leaching contain zinc ferrite, entrained zinc sulfate, undissolved zinc oxide, and other values, such as lead, copper, silver and gold. Although hot acid leaching improves copper and zinc recovery, the process was not widely used in the past because of the difficulty of separating iron which forms a gelatinous precipitate. Usually the residue was treated pyrometallurgically and often still is.

In addition, three commercially successful methods have been developed in which iron(III) precipitates are formed that filter and wash well, and which increase the recovery of zinc by 10–15%.

The jarosite process separates iron(III) from zinc in acid solution by precipitation of $MFe_3(OH)(SO_4)_2$, where M is an alkali metal (usually sodium) or ammonium (see Fig. 2) (40,41). Other monovalent and hydronium ions also form jarosites which are found in the precipitate to some degree. Properly seeded, the relatively coarse jarosite can be separated from the zinc-bearing solution efficiently. The reaction is usually carried out at 95°C by adding ammonia or sodium hydroxide after the pH has been adjusted with calcine and the iron oxidized. The neutral leach residue is leached in hot acid (spent + makeup) with final acidity >20 g/L and essentially all the zinc, including ferrite, is solubilized. Ammonium jarosite is then precipitated in the presence of the residue or after separating it. If the residue contains appreciable lead or silver, they are first separated to avoid loss to the jarosite waste solids. Minimum use of calcine in jarosite neutralization is required for maximum recovery of lead and silver as well as zinc and other metals.

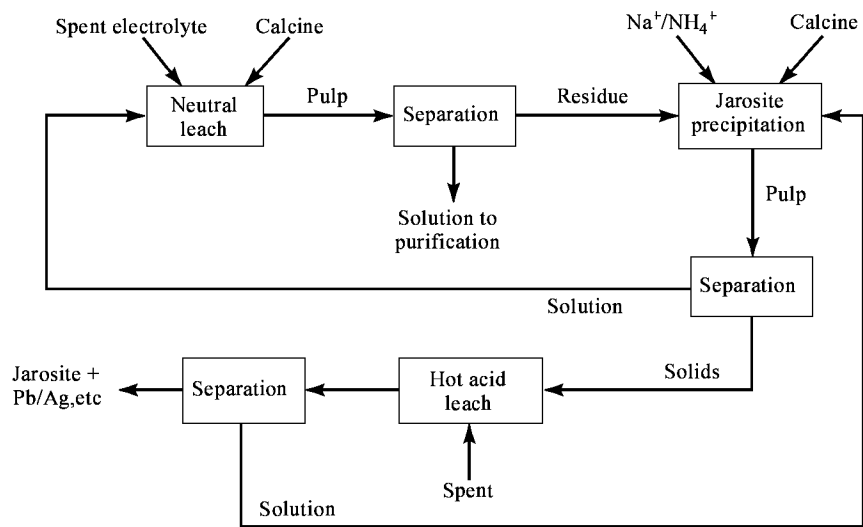


Fig. 2. Flow sheet for modified jarosite precipitation process.

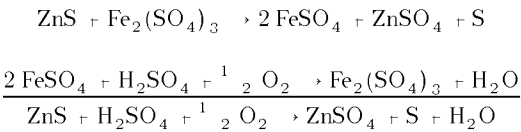
In a modification, the conversion process, the jarosite residue is hydrothermally decomposed to hematite by autoclaving at 220–250°C. This solubilizes zinc and other metal values and the hematite has a potential for iron recovery. Hematite stockpiles are less of a problem than jarosite because hematite is denser and holds up less of the soluble metals.

The jarosite process controls both alkali metals and sulfate in the zinc-plant circuit and consumes little neutralizing agent. It is used in 16 plants worldwide, which account for ca 25% of the noncommunist world's primary zinc output.

The goethite process precipitates crystalline αFeO·OH (goethite) as well as βFeO·OH, αFe₂O₃, and amorphous phases. The reaction is carried out at 90°C and pH 3.0, for 4–6 h in either batch or continuous fashion, and the iron(III) ion must be kept <1 g/L. Both jarosite and goethite solids are usually lagooned.

The hematite process is used at the Iijima plant of the Akita Zinc Co. in Japan, where it started in 1972, and at Ruhr-Zink, Datteln, FRG. In the Iijima process, neutral leach residue is leached with hot acid and the iron(III) in the leachate is reduced with sulfur dioxide at ca 100°C. Copper is precipitated with hydrogen sulfide although it is not essential to the process. Ruhr-Zink reduces with concentrate, zinc sulfide, giving elemental sulfur as product. Sulfur and unreacted zinc concentrate are coagulated at elevated pressure and temperature and separated by classification from finer lead–silver residue. The solution is neutralized to pH 4.5 with limestone to yield gypsum which is filtered and sold. This leaves iron(II) sulfate solution from which αFe₂O₃ is precipitated by oxidizing at ca 190°C under an oxygen pressure of 8.4 kPa (63 mm Hg). The hematite product is denser than jarosite or goethite; it can be sold to steel and cement plants.

Direct Leaching of Concentrates. Sherritt Gordon Mines, Ltd., has adapted the process first used on nickel sulfide ores to zinc sulfide oxidation with air in aqueous slurry under pressure (42,43). The concentrates are leached directly with return acid from the cells and the sulfide is converted to free sulfur:



Thus, roasting is avoided. The process, especially amenable to high iron and copper concentrates, has been installed by Cominco, Ltd. (44) at Trail, B.C., Canada, and will be installed at the Kidd Creek Mines, Ltd., plant at Timmins, Ontario.

Purification. Purification of the electrolyte is extremely important since the presence of impurities, even in trace amounts, can seriously impede or halt production (45,46). Impurities may lower the hydrogen overvoltage causing the electrical energy to be consumed in depositing hydrogen instead of zinc. The reversible deposition potential of hydrogen from acid onto platinum is 1.70 V but on zinc, because of overvoltage, it is 2.4 V. The standard potential needed for zinc reduction is 2.35 V (45). Therefore, if the hydrogen potential is lowered by a mere 0.05 V, it is deposited strongly in competition with zinc. Very low impurity levels are usually very harmful and some exhibit synergism. For instance, although cobalt is tolerated in some cells at up to 10 mg/L, it increases the damage caused by germanium. Therefore, it is difficult to assess the true effect of various impurities especially since temperature, current density, and electrolyte concentration are also factors. The effects of various impurities are given in Table 5.

Table 5. Effect of Impurities in Zinc Electrowinning

Impurity	Reported range—11 plants, mg/L	Class ^a	Effect			
			Current efficiency	Resolu-tion	Deposit	Other
germanium	0.005–0.2	4	lowers	high	spongy	worse with cobalt
tellurium	<0.001 ^b	4	lowers	high	uneven	worse with cobalt
selenium	<0.002 ^b	4	lowers	high	uneven	
arsenic	0.003–0.02	4	lowers	high	corrugated	worse with cobalt
antimony	0.01–0.03	4	lowers	high	beadypoor ad-hesion	worse with cobalt or germanium
copper	0.05–0.2	4	lowers	yes		
nickel	<0.01–0.5	3	lowers	mild	holes	
cobalt	0.03–2.0	3	lowers	mild	holes	reduces lead deposition
tin	<0.02 ^b	3	lowers	yes	filmy	
iron	0.2–25	3	lowers			may reduce anode corrosion
cadmium	0.01–5	2				deposits with zinc
lead	1 ^b	2				deposits with zinc
thallium	0.5–5 ^b					increases lead depo-sition, deposits
aluminum	10 ^b	1				increases electrolyte resistivity
magnesium	6.5–12 g/L	1				increases electrolyte resistivity

manganese	3–3.5 g L	1	oxide forms on anode, lowers corrosion, may insulate anode
chloride	20–100		corrodes lead anodes
fluoride	2 ^b		corrodes electrodes, causes sticking of zinc to cathode

^a Class 1. Higher reduction potential than zinc. Will not deposit on cathode but will affect conductivity. 2. Lower reduction potential than zinc. Hydrogen overvoltage above 0.65 V. Deposits with zinc but does not affect zinc deposition. 3. Reduction potential between zinc and hydrogen. May deposit and redissolve, affecting current efficiency. Causes local lowering of hydrogen overvoltage, making pits, holes, etc. 4. Reduction potential below hydrogen. Hydrogen overvoltage below 0.65 V. Deposits and causes hydrogen evolution. Tends to form hydrides.

^b One plant reporting.

Purification actually starts with the precipitation of the hydrous oxides of iron, alumina, silica, and tin which carry along arsenic, antimony, and, to some extent, germanium. Lead and silver sulfates coprecipitate but lead is reintroduced into the electrolyte by anode corrosion, as is aluminum from the cathodes; and copper by bus-bar corrosion.

Purification is primarily based upon displacement from solution, ie, so-called cementation with metallic zinc of the metals below zinc on the electromotive series. For this purpose, zinc is melted and atomized with an air jet to so-called zinc dust, which is screened into fractions in the range 70–300 μm (ca 50 to 200 mesh) for use. Zinc is added in large excess over stoichiometric and about 5° of the cathode zinc is recycled for this purpose. Cobalt is removed in some plants by precipitation with α-nitroso-β-naphthol. In one plant, dimethylglyoxime is used to insolubilize nickel. Most importantly, magnesium, alkali metals, chloride, and fluoride are not removed by iron precipitation or cementation but by stripping the zinc from a portion of the electrolyte and discarding the solution. The plant solutions are saturated with calcium sulfate and incoming calcium is precipitated as gypsum at relatively cool points, eg, the cooling towers and coils. Gypsum is an important outlet for sulfate which enters with the calcine in small amounts.

Purification is highly specific for any given plant and ore composition. In general, the zinc-dust particle size, pH, temperature, agitation and presence of additives (such as copper, arsenic, and antimony) are important in determining the reaction rate. Both batch and continuous procedures are employed. Final polishing stages using minimal zinc are sometimes added. The three most popular purification flow sheets are described in Table 6 (47), but many variations are practiced, eg, new three-stage procedures (48,49).

Table 6. Purification of Zinc Liquor

Stage	Conventional arsenic ^a	Conventional antimony ^a	Reverse antimony ^b
stage 1			
remove	Co, Cu, Ni, As, Sb	Co, Cu, Ni, Cd	Cu, Cd, Tl
reagent	coarse Zn, CuSO ₄ ^c , As ₂ O ₃	coarse Zn, Sb ^d	Zn
temperature, °C	90	65–75	ambient
pH	4.0		
stage 2			
remove	Cd, Tl	Co, Cu, Ni, Cd	Co, Ge, Ni
reagent	Zn	fine Zn	Zn, Sb
temperature, °C	70–80	ambient	90
pH	3.0		

^a To remove Co and Ni in stage 1 (hot).
^b Reverse antimony: stage 2, Sb (hot) for Co, Ni removal.
^c CuSO₄ added if Cu is <400 mg L.
^d α-Nitroso-β-naphthol may be substituted for Sb to precipitate Co.

Copper sulfate, in small amounts, activates the zinc dust by forming zinc–copper couples. Arsenic(III) and antimony(III) oxides are used to remove cobalt and nickel; they activate the zinc and form intermetallic compounds such as CoAs (49). Antimony is less toxic than arsenic and its hydride, stibine, is less stable than arsine and does not form as readily. Hydrogen, formed in the purification tanks, may give these hydrides and venting and surveillance is mandatory. The reverse antimony procedure gives a good separation of cadmium and cobalt.

Aeration must be avoided since it can oxidize and resolubilize the cemented (precipitated) impurities. Filter presses are used after each step and the cakes are leached to recover various values. For example, cadmium is dissolved, recemented with zinc, and recovered on site either electrolytically or by distillation. A copper residue of 25–60° copper is sold for recovery elsewhere. The other impurities cannot be recovered economically with the exception of cobalt in some plants.

Electrolysis. The net electrochemical reaction is ZnSO₄ + H₂O → Zn + H₂SO₄ + 0.5 O₂. Electrolysis is carried out in cells with 17–44 aluminum cathodes and 18–45 lead anodes per cell (30,50). Cell tanks are usually concrete lined with lead, rubber, or plastic. Anodes and cathodes in a given cell are connected in parallel with an imposed voltage of 3.3–3.8 V which is higher than the standard 2.35 V because of the resistance of the electrolyte, gaseous, and solid films on the electrodes, and contact resistance. Energy consumed is 3.3 kW·h kg zinc at slightly over 50° efficiency. Cells are arranged in groups connected in series with direct current supplied by rectifiers which are controlled by load-optimizing equipment in many plants. Thus, the power consumption varies, lower during peak demand period and higher during off-hours.

Increasing the current density increases the zinc-output rate per cell. Electrolytic plants were formerly classified as either low current density (200–450 A m²)–low acid (<6%) or high current density (1000 A m²)–high acid (22–28°). The ratio of zinc to acid in solution must be kept high enough to avoid resolution of cathode zinc and therefore high acid plants operate at higher zinc levels. At higher current density, more heat must be removed and solution purity is more critical but the return acid can more readily dissolve zinc ferrite formed in roasting. The choice of current density to use is therefore determined by many factors. Modern plants have current densities of 300–600 A m² and achieve efficiencies of slightly over 90° in good operation (see Table 7). Glue (<100 mg L) improves the smoothness of the cathode deposit and moderates the effect of impurities (52).

Table 7. Cellhouse Operating Parameters^a

Parameters	Big River Zinc, Sauget, Illinois	Pasminco Ltd., Tasmania	Canadian Elec-trolytic Zinc Valleyfield, Quebec	Union Miniere Balén, Belgium ^b	Savage Zinc Clarksville, Tennessee	Mitsubishi Akita, Japan	Mitsui Hikoshima, Japan
electrolyte							

Zn, g L	55	46	70	50	90 (65) ^c	47	na
H ₂ SO ₄ , g L	200	100	200	190	128 (175) ^c	115	na
temperature, °C		35	35	30	32	36	36
cell arrangement	cascade	row ^d	row	row	row	row	row
cathodes cell	27	48	36	44	49	39	33
cathode area, m ²	0.65	0.65	1.1	2.6	2.6	1.6	1.6
current density, A m ²	807	540	500–700	300–400	375	370–430	300–600
voltage	3.8	3.5	3.5	3.3–3.5	3.3	3.55	
liner	lead	lead	lead	PVC	PVC	rubber	rubber
deposition, h	24	72	24				
stripping type	manual	manual	manual ^e	auto knives	auto knives	auto hinged rubber edge-knives	auto high frequency shake
removal cycle, h				48	48	48	32–48

^a Older plants are described in ref. 51.

^b The Zinc Corporation of America plant at Bartlesville is, like the Savage Zinc Co. plant, based on the Union Miniere design.

^c Secondary cells.

^d New section.

^e Old section.

The purified zinc solution is fed to the electrolyte recirculating stream at a rate that holds the composition of the electrolyte constant, commonly ranging from 100–200 g H₂SO₄ and 45–70 g Zn L. Continuous monitoring of the density and conductivity of the spent acid aids in control. This is important since the range of acidity is narrow for maximum current efficiency at any given current density.

The cells are fed individually but cascading is practiced in some older plants. In this system, electrolyte overflows from one cell to the next in a series of 3–9 cells. More commonly, cells are placed in rows, with each cell overflowing into a common spent-acid launder.

Cooling by means of evaporative cooling towers is required to maintain a constant temperature of 30–40°C. At higher temperatures, the deposit is rougher, impurity effects are more pronounced, lead codeposition is favored, and the manganese dioxide formed at the anode increases and tends to adhere rather than fall to the bottom of the cell.

Oxygen evolved from the anodes as well as some hydrogen from the cathodes produces a mist which is trapped by a froth maintained by adding cresylic acid, sodium silicate, and gum arabic, or glue plus cresol. Alkaline-earth carbonates prevent lead contamination of the cathode zinc. Most of the lead is deposited in the cell sludge as insoluble carbonate–sulfate.

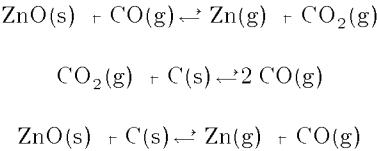
Cells and anodes are cleaned on a cycle of 40–90 d when the sludge, mostly manganese dioxide, is pumped out. The manganese dioxide coating on the anodes is removed with water jets and the anodes mechanically straightened. Aluminum cathodes are welded to aluminum header bars which have copper contact points welded or cast on. The aluminum is polished at a frequency depending upon the corrosiveness of the electrolyte. Cathodes are often anodized to minimize corrosion by brief contact with the "live" anodes upon re-entering the cell after stripping. Chlorides and fluorides are corrosive and make the zinc adhere to the cathode. This is especially critical in mechanized stripping operations which are rapidly replacing the manual method. Antimony chemicals are added in some plants (<0.08 mg Sb L) to combat sticking. Manual stripping of the deposit from the cathodes has always been an expensive and onerous task. Several plants adapted automatic mechanical strippers to existing cellhouses and, in 1969, Union Minere started a fully automated cellhouse at Balen, Belgium, featuring jumbo cathodes of large area (2.6 m²).

Melting, Casting, Zinc Dust. Cathode zinc is melted, usually by induction heating, and distributed by pump to a casting machine which produces 26-kg slabs of special high grade zinc. Larger castings of 1090 kg and 450 kg, for instance, are made in preheated molds. Some of the zinc may be pumped to an alloying furnace for the addition of metals such as lead, aluminum, and copper. These alloys are cast into slabs, jumbo blocks, and special shapes.

Some of the melted zinc is fed to the zinc-dust unit where the molten zinc may be dropped from a crucible through a small orifice (2.5 mm) to be atomized in a blast of air. Solidified droplets are collected in a chamber and screened to the proper size for purification and cadmium plant cementation. Frequently, coarse (+70–200 μm) and fine (–70 μm) fractions are required.

Ammonium chloride is used as a flux in the melting furnace because the large surface of the cathodes favors the formation of dross, ie, oxide-coated globules of zinc. The dross is separated by liquation or air-swept milling into metal and oxide fractions. In the latter, the oxide fraction is swept out of the mill and can be returned to roasting for the elimination of chloride. Metallic zinc is recycled. Overall melting efficiency is 96–98%.

Pyrometallurgical Processes. Zinc pyrometallurgy is based upon reduction of zinc oxide (53).



The lowest temperature for the reaction is 857°C but 1100–1300°C is required for acceptable rates. In this range, the two component reactions proceed at about the same rate and the reduction is diffusion controlled (54). Both reactions are reversible and the overall reaction is endothermic, requiring about 5.5 GJ t Zn (1.2 × 10⁹ cal short ton Zn) at 1200 K.

Carbon monoxide and dioxide oxidize zinc vapor below 1100–1300°C although only the carbon dioxide reaction is significant. Rapid condensation of the zinc vapor avoids the formation of zinc-oxide-coated droplets, so-called blue powder.

At first, batchwise horizontal retorts were used for smelting, and later continuous vertical retorts, both externally fired. Continuous, internally heated furnaces such as the electrothermic furnace followed, and the last important development was the Imperial Smelting blast furnace.

Imperial Smelting Furnace. In the ISF process, zinc vapor is produced in a blast furnace along with lead. Although air is introduced to provide heat through the burning of coke, the atmosphere in the upper portion of the furnace is not oxidizing despite the carbon dioxide content (CO₂ CO 10% 25%) because of the high temperature, 1000–1050°C. To prevent reaction of carbon dioxide with zinc vapor, the gases leaving the furnace are rapidly cooled to 550–600°C in splash condensers (see Fig. 3).

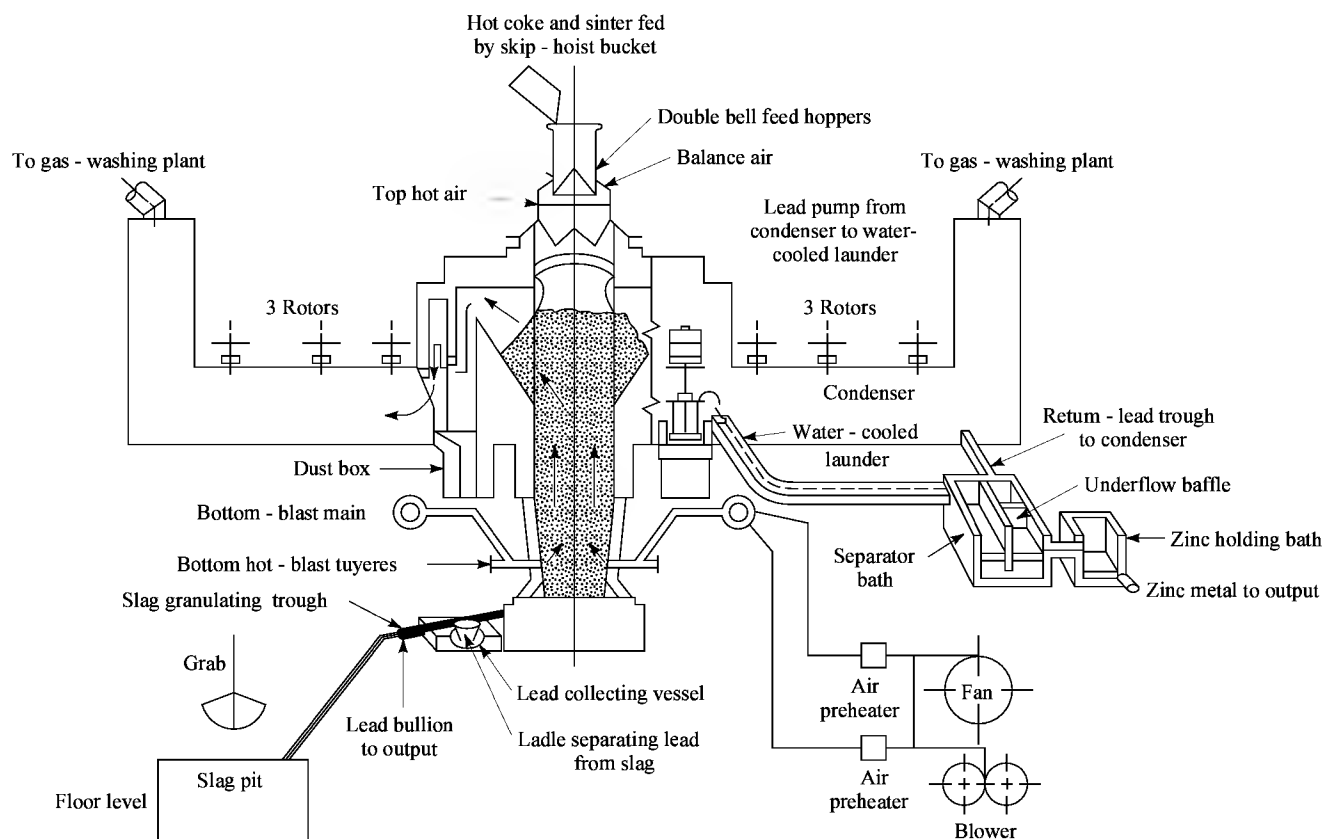


Fig. 3. Imperial Smelting furnace.

The Imperial Smelting Corporation, Ltd., brought its blast furnace on-stream at Avonmouth, UK, in 1952 (55), and in 1995 there were 12 operating ISF plants around the world ranging in capacity from 30,000 to 105,000 metric tons per year. There are none in the United States, Mexico, or Canada since the Brunswick Mining & Smelting Corporation plant at Belledune, N.B., Canada, was converted to straight lead in 1972. The ISF can handle complex zinc-lead-copper ores, zinc-containing dust from steelmaking, drosses, leach plant residues, etc, in a large, economical unit. The ISF is more energy-efficient (37.9 GJ t or 32.6 × 10⁶ Btu short ton Prime Western Zinc) than either the vertical retort (60.2 GJ t or 51.8 × 10⁶ Btu short ton) or the electrothermic (72.2 GJ t or 62.1 × 10⁶ Btu short ton) processes (56). In fact, the figures favor the ISF over electrolytic (49.8 GJ t or 42.9 Btu short ton) for Prime Western and 44.8 vs 50.1 GJ t (38.6 vs 43.1 × 10⁶ Btu t) for special high grade (57). The electrolytic process, however, uses lower cost fuel in some locations and has other cost advantages including labor and environmental.

Feeds have lead-zinc ratios from 0.45 to 0.82 (58). The feed mix, primarily sulfide concentrate, is updraft-sintered and oxidized with the off-gases passing to an acid plant. Most of the cadmium is eliminated in the sintering step and the lead oxide, if above 7% lead, acts as a flux to produce the hard sinter required for the blast furnace (58). Only a small portion of the sintering heat is recovered compared to the waste heat recovered in fluid-bed roasting. Hot, briquetted, roasted concentrate or oxidized ore is an alternative to sintering, especially where lead is low. This approach is being used to supplement sintered feed up to 25%. Successful trials of 100% briquetted feed have been made, opening the possibility of using the ISF for zinc alone.

Sinter and metallurgical coke are preheated to 400 and 800°C, respectively, and fed into the top of the blast furnace through a double-bell hopper. Preheated air (800–1000°C) is blown in through tuyeres near the bottom providing heat and carbon monoxide by oxidation of the coke. Lead, silver, and copper are deposited in the molten lead bullion tapped from the furnace bottom. Slag floats on top of the bullion and is tapped periodically. Zinc vapor exiting the top is heated to 1000°C by combusting the off-gas with injected air and is rapidly condensed in a spray of molten lead after which the molten metals flow to a cooler where a zinc-rich layer overflows and the high lead bottom recirculates to the splash condenser. Zinc recoveries range from 92 to 95% and, with lead at 1.1–1.3%, the zinc is in the Prime Western class.

Since a large amount of heat must be absorbed at the condenser, the lead recirculation rate is very high, 300–500 t h per condenser or 400 t lead per ton zinc. Heat lost to the cooling water is not recovered but heat is recovered from the condenser off-gases (LCV or low calorific value gas) at 450°C in most plants to preheat air, coke, etc. Consideration is being given to using a waste-heat boiler at this point. The Hachinohe smelter in Japan has reported on heat recovery from LCV gas and other energy savings (59).

Grades purer than the Prime Western product have been made by vacuum dezincing the liquid alloy to recover 99.9% zinc at 0.02% lead. Distillation (New Jersey Zinc process) produces a purer product at a considerable cost in energy.

The ISFs range in shaft area from 15.3 to 27.1 m² and produce from 112 to 334 t d. The standard furnace is 17.2 m² and its typical performance is given in Table 8 (58).

Table 8. Imperial Smelting Furnace Performance^a

Characteristic	Value
production ^b , t d	
Zn	256
Pb	128
input ratio	
Pb:Zn	0.5
C:Zn	0.75
Zn in slag, %	7.0
slag:slab Zn	0.65
recovery, %	
Zn	94.5
Pb	95.0
carbon burned per metric ton slab Zn	0.82

^a Ref. 58.

^b Full blast.

- Efficiency is improved and cost lowered by the following changes:
- Hot (500–700°C) binderless briquetting of oxidized zinc and lead feed materials to replace sintering (see SIZE ENLARGEMENT).
- Changes in furnace design. More uniform charge distribution, new furnace shapes, and tuyere design.
- Drying of the air blast lowers coke usage 3% for each 1% of water vapor removed.
- Heat recovery from furnace off-gas (LCV gas) normally used to preheat coke and blast air; LCV gas has also been burned to heat melting baths and generate steam and power.

Vertical Retort. The vertical retort process was developed by The New Jersey Zinc Company (60). Production started in 1929 at Palmerton, Pennsylvania. A charge of 25% bituminous coal, 10% anthracite, 10% recirculated fines, and 55% sinter is mixed with suitable binders and densified in a chaser mill. The mix is briquetted and coked autogenously, that is, the volatile from the coking process burn in the bed of briquettes to supply the needed heat. Briquettes leave the coker at ca 700°C and are transferred to the top of the retort where they are charged on a periodic cycle. Radiant heat from the walls drives the reduction as the briquettes descend. The loaf-shaped briquettes, strong enough to withstand breakage, are extracted from the bottom through a water seal. Carbon monoxide forces the zinc vapor up the column; ca 95% of the metal is condensed and the balance scrubbed out.

The Palmerton, Pennsylvania, plant had 43 retorts with an output of ca 8 t/d per retort. Recovery was ca 94% when the plant was shut down in 1980. The zinc contained ca 0.3% lead, 0.10% cadmium, and 0.01% iron plus minor impurities. Lead and aluminum are added to produce galvanizer's zinc.

Zinc of 99.995% purity is produced in fractional-distillation columns which were developed along with the vertical retorts. Refining is a two-stage procedure in which higher boiling lead and iron are separated in the first column, whereas lower boiling cadmium is distilled in a second column (61).

Electrothermic Process. Electric-arc furnaces with molten iron-slag baths have never been successful despite many attempts (62). However, a process developed by the St. Joseph Lead Company (now Doe Run Company), based upon resistance heating, was employed in four plants in 1981 (34). There is relatively little dusting and zinc condensation is efficient.

Sintering in two stages the second with sand, produces a high purity, hard sinter which is required for the oxide and high grade zinc furnaces. Single-stage sinter is fed to the Prime Western and intermediate-grade circuits. Both sinters are sized to 9–25 mm. Coke is fed at ca 300%. Small furnaces have daily outputs of 16–25 t, large ones 100 t. Current consumption is 1250 kW per electrode pair at 200–230 V. Vapor exits through an annular vapor ring and is led to the condenser.

The Weaton-Najarian zinc condenser was commercialized in 1936. The condenser and cooling well of the electrothermic furnace hold 48 t of molten zinc. Hot zinc-laden gases bubble through the zinc in the condenser and cause rapid circulation through the cooling well which is kept at 480–500°C by water coils. The off-gases are scrubbed and burned for fuel value. Scrubber water is ponded to recover blue powder. *Horizontal Retort.* In 1800, the first commercial zinc process made use of the horizontal retort. In 1980, only three such plants remain because they are not competitive in terms of labor and fuel costs. Furthermore, the dust produced presents a serious pollution problem. Nevertheless, in 1956, the tonnage of zinc produced from horizontal retorts was above that of any previous year. The only remaining operation is in Russia with a capacity of 10,000 annual MT.

Banks of Belgian horizontal cylindrical retorts are arranged in a furnace containing 500–700 retorts. Heat is applied externally by natural or producer gas. The retorts are charged with a damp mixture (35–40% zinc) of sinter, blue powder, and carbonaceous reducing agent (50%). The cycle is ca 48 h, and 40 kg zinc is produced per retort. The zinc vapor is liquefied in a clay condenser. Metal is drawn periodically in 10-kg lots into a ladle which is skimmed to remove blue powder and impurities. Slab zinc is cast in molds directly from the ladle and the casting is skimmed. Recoveries are 88–95%, and the quality of the metal depends upon the composition of the charge.

Secondary Recovery. Zinc is recovered as metal, dust, and chemicals (including oxide) from secondary sources, mostly scrap (see also RECYCLING). So-called old scrap originates from die castings and engraver's plates, whereas new scrap, such as drosses, skimmings, flue dust, clippings, and residues, originates in various processes. The amount of old scrap has remained at ca 5–6% of total production for the last two decades but it may be possible to increase recovery from die castings and steelmaking dust. In 1980, secondary slab zinc was produced in the United States at nine plants as well as by two primary producers. Zinc dust was produced at eight plants, mostly from secondary materials. Recovery practices vary greatly, but zinc products are usually made from metallic scrap by melting and distillation. Dust is made by rapid condensation of zinc vapor in oxygen-free atmospheres; the spherical particles have diameters of 1–20 μm (2).

A novel process, called Zincex, was brought on-stream in 1976 by Metalquímica del Nervion in Bilbao, Spain, where 8000 t/yr is produced from pyrite cinders (63). The cinders are given a chloridizing roast and leached to produce a solution containing 20–30 g/L zinc, 18–25 g/L iron, 60–100 g/L chloride, and 120–155 g/L sulfate. Solvent extraction in two stages recovers 98% of the zinc. The first stage uses a secondary amine to extract ZnCl₄²⁻ which is stripped with water. Di-2-ethylhexyl phosphoric acid absorbs zinc ion in the second stage where spent electrolyte is the stripping agent. Thus, zinc is well-separated from chloride, iron, and other harmful impurities.

Economic Aspects

The United States, the former USSR, Canada, and Australia have the largest known reserves of zinc ore which should permit mining at current levels into the next century (see Tables 2 and 3). World mine production of recoverable zinc between 1970 and 1996 is given in Table 9 (64–66). Mine production in the United States in the 1970s was lower than in the 1960s, reached a minimum in the mid-1980s and has risen since then. The U.S. share of world production has historically been 8–9%.

Table 9. United States, Canada, North America, South America, and World Mine Production, 1000 t

Year	United States	Canada	North America ^a	South America ^b	World	World United States ratio
1970	484.6	1239.2	1903.6	397.3	5464.0	11.28
1975	425.8	1055.2	1738.4	544.2	5988.1	14.06
1977	448.3	1300.2	2052.8	635.2	6608.2	14.74
1979	293.8	1204.4	1766.4	643.2	6335.1	21.56
1980	343.9	1058.7	1661.1	649.1	6214.1	18.07
1996	603	1234	2248	1101	7223	11.97

^a United States, Canada, Mexico, Guatemala, Honduras, Nicaragua.

^b Argentina, Bolivia, Brazil, Peru, Chile, Colombia, Ecuador.

Alaska was the principal zinc-producing state during 1992, followed by, in order of output, Tennessee, New York, Missouri, Colorado, and Montana.

Zinc consumption is broken down into slab and all-classes classifications. The latter includes zinc-containing materials made from ores and residues as well as from secondary (scrap) zinc. Consumption and production figures of slab zinc show that a substantial difference exists which is made up by imports (see Table 10). The imports consist of ore and concentrates based on zinc content, slab zinc in the form of blocks and pigs, zinc dust by gross weight, waste and scrap on a gross weight basis, and dross and skimmings based on zinc content. In 1970, ore and concentrate imports were about equal

to U.S. mine production and slab zinc imports represented 30^o of U.S. production. In the late 1970s, imports of ore and concentrates decreased below U.S. production and imports of slab zinc increased to exceed U.S. production from 1976 onward. Imports of ore and concentrates in 1996 came from Peru (53^o o) and Mexico (47^o o). Canada supplied ca 61^o o of slab zinc and Mexico and Spain each supplied about 11^o o. For 1996, the figures for U.S. exports are given in Table 11.

Table 10. U.S. Production, Consumption, and Prices of Slab Zinc, 1000 t

Use	1970	1976	1978	1980	1996 ^d
production					
domestic ores, recoverable Zn content	484.6	439.6	302.7	343.9	603
slab zinc					
from domestic ores	366.5	346.4	267.4	234.4	178
from foreign ores	429.9	106.1	139.4	113.7	40
from scrap	70.0	62.2	34.8	28.8	148
<i>Total</i>	<i>866.4</i>	<i>514.7</i>	<i>441.6</i>	<i>376.9</i>	<i>341</i>
imports					
ores and concentrates (Zn content)	477.0	88.1	187.4	129.9	15
slab zinc	245.3	648.2	617.8	410.6	827
other ^a	9.9	18.4	19.6	11.5	2
consumption					
slab zinc	1076.8	1135.8	1063.3	817.9	1205
all classes	1420.6	1479.0	1449.2	1153.6	1297
price ^b , Prime Western, yearly average, ¢/kg					
United States	33.778	81.607	68.291	82.529	
London (LME) ^c	29.035	71.230	59.248	76.033	102.289 ^e

^d Ref. 66.

^a Zinc dust, gross weight; waste and scrap, gross weight; dross and skimmings zinc content.

^b Ref. 67.

^c Ref. 68.

^e Special High Grade.

Table 11. U.S. Zinc Exports, 1996

Material	1000 t
ores and concentrates	425
zinc pigs and slabs	
zinc sheet and strip	6.6
dross, scrap, ashes, and skimmings	58.3
<i>Total</i>	<i>490.9^a</i>

^a Refs. 11 and 66.

Production processes are given in Table 12. Electrolytic processes are dominant because of lower cost and fewer environmental problems. Production of slab zinc in Tennessee commenced in 1978 when Jersey Miniere Zinc Company began operation of its new 90,000 t/yr electrolytic smelter at Clarksville. Today, Tennessee is the leading producer of slab zinc. Although the U.S. demand for zinc metal in the past 16 years has increased by 47^o o, smelting capacity has declined by almost 50^o o. Plants closed because they were obsolete and could not meet environmental standards or obtain sufficient concentrate. Consequently, slab zinc has replaced concentrates as the principal import form. This situation is expected to prevail up to the year 2000 (69–71).

Table 12. Distribution of Zinc Production Processes

Process	Date of commercialization	Worldwide production capacity, 1996, 1000 MT ^a
electrolytic	1915	6,456
horizontal retort	1800	10
vertical retort	1930	230
electrothermic	1936	416
Imperial Smelting (ISF)	1952	863

^a Ref. 66.

Electrolytic and distillation processes are used to produce primary slab zinc at smelters from ores and concentrates, whereas redistillation is used to recover zinc from secondary zinc materials at both primary and secondary smelters (see Table 13) (64,65). In 1965, nearly 60^o o of the primary slab zinc was manufactured by the distillation process. However, in 1975, electrolytic slab zinc became dominant and by 1996 its share increased to 78^o o.

Table 13. Slab-Zinc Production in the United States by Various Reduction Methods, 1000 t^a

Year	<i>Method of reduction</i>				
	Primary		Secondary redistilled		Grand total
	Electrolytic	Distilled	At primary smelters	At secondary smelters	
1970	356.8	439.6	59.7	10.3	866.4
1975	211.7	185.8	31.7	20.8	450.0

1977	210.5	197.9	26.4	19.5	454.3
1980	295.2	52.9	15.3	13.4	376.8
1996	170	48	98	50	366

^a Refs. 64–66.

In 1996, secondary production was 40^o of primary production, whereas in 1980 it was 89^o of primary zinc production. The tonnage of secondary zinc has increased largely because of recovery of zinc from electric arc furnace dust from steel making. The balance of the zinc-containing scrap utilized in 1980 went into the production of zinc dust (9^o o); zinc, magnesium, and aluminum alloys (3^o o); brass and bronze alloys (59^o o); and pigments and chemicals (19^o o). As a result of the closing of U.S. primary production, greater reliance will have to be placed on secondary zinc and imports (69–71).

The world production of mined zinc ore and of zinc-slab production are given in Table 14. In 1996, the 13 producing countries listed accounted for 84^o of the zinc mined in the world. In many cases, ore is smelted in the country in which it is mined, but an active market in zinc concentrates exists. The most notable exceptions are Germany, France, Belgium, Italy, and the Netherlands, which produce metal from imported ore in varying degrees. Mining countries that have little or no smelting capacity are Peru, Ireland, and Sweden.

Table 14. World Mine Production^a and Slab-Zinc Production, 1000 t^b

Country	1976		1978		1980		1996	
	Mine	Slab zinc	Mine	Slab zinc	Mine	Slab zinc	Mine	Slab zinc
	pro-duction		pro-duction		pro-duction		pro-duction	
Australia	461.9	249.2	473.3	294.3	470.6	306.0	1016	327
Belgium		244.4		232.9		238.0		207
Canada	1145.0	472.3	1245.3	495.4	1058.7	586.7	1234	716
France		233.5		231.2		252.8		324
Germany		304.8		306.8		368.8		327
Ireland	62.8		176.0		228.8		163	
Italy		191.2		177.6		206.4	12	269
Japan	260.0	742.1	274.6	767.9	236.9	743.9	80	599
Mexico	259.2		244.9		239.3		374	223
Netherlands		123.2		135.3		168.1		207
People's Republic of China	135.0	150.0	150.8	160.0	155.0	160.6	1040	1079
Peru	413.7		457.5		492.0		758	175
Poland	215.0	237.0	231.0	222.0	237.2	215.0	153	166
Spain	82.3		143.2		170.4		140	361
Sweden	128.3		167.4		170.0		161	
United States	483.0	514.7	337.0	441.5	343.9	358.6	603	366
USSR ^c	1020.1	1000.0	1030.0	1055.0	1020.1	1088.6	305 ^d	400 ^d
other	1586.2	1327.1	1509.7	1521.3	1391.3	1502.4	1184	1548
wodd	6252.5	5789.5	6440.7	6041.2	6214.2	6195.9	7223	7294

^a Based on zinc content.

^b Refs. 64,65.

^c Estimated.

^d Former USSR countries.

Zinc consumption is categorized in five semifabricating markets (see Table 15). Galvanizing was the main market for zinc in the 1970s followed by zinc-base casting alloys and brass and bronze. Depressed construction and automotive industries caused a decline from 1979 to 1980 of ca 18^o o, and the die-casting business declined 34^o o and galvanizing 24^o o.

Table 15. Distribution of U.S. Slab-Zinc Consumption, 1000 t^a

Use	1970	1976	1978	1980	1996 ^c
galvanizing					
sheet and strip	229.7	223.8	268.7	218.9	
wire and wire rope	28.0	24.3	22.8	20.6	
tube and pipe	58.5	44.4	47.4	34.8	
fittings (for tube and pipe)	8.6	5.9	6.9	5.3	
tanks and containers	3.6	3.0	2.9	3.8	
structural shapes	17.0	31.4	33.3	19.4	
fasteners	4.8	3.7	4.8	2.7	
pole-line hardware	9.0	4.3	4.9	3.8	
fencing, wire cloth, and netting	16.4	20.0	25.0	12.1	
other and unspecified uses	54.6	32.2	37.4	23.6	
Total	430.2	393.0	454.1	345.0	654
zinc-base alloys					
die-casting alloys	411.4	380.8	346.0	202.6	
dies and rod alloys	0.8	0.9	0.5	0.2	
slush and sandcasting alloys	9.1	5.7	7.6	5.6	
Total	421.3	387.4	354.1	208.4	229
brass and bronze ^b					
sheet, strip, and plate	55.9	82.7	70.2	37.6	
rod and wire	37.6	49.5	46.3	32.4	
tube	8.2	6.7	6.8	4.6	
castings and billets	4.2	3.8	4.4	1.8	
copper-base ingots	9.0	7.0	6.6	17.0	
other copper-base products	0.9	1.1	7.2	3.0	
Total	115.8	150.8	141.5	96.4	155

rolled zinc	37.2	27.1	24.9	21.1	^d
zinc oxide	39.8	35.4	37.2	29.9	80
other ^c	33.0	35.3	38.9	25.0	92
estimated undistributed consumption				91.2	
<i>Grand total</i>	<i>1077.3</i>	<i>1029.0</i>	<i>1050.7</i>	<i>817.0</i>	<i>1210</i>

^a Refs. 64, 65, 72.

^e Ref. 66.

^b For zinc in brass objects, see Copper alloys.

^d Included in other.

^c Zinc dust, light metal alloys, desilvering lead, bronze and brass powders, zinc chemicals.

Die-Casting Alloys. Consumption of die-casting alloys increased rapidly in the 1940s because of usage in the automotive industry. U.S. zinc tonnage used in die castings reached a peak in 1965 with 578,766 metric tons. The decline in the zinc die-casting market can be attributed to reduced automotive sales and competition from aluminum, magnesium, and plastics, particularly where weight limitations and reductions are required. However, the development of thin-walled die castings and improved alloys and finishing techniques indicates that die casting will continue to consume large quantities of zinc in the transportation and appliance industries. The price of zinc relative to that of aluminum and plastics and public preference for metal parts because of their durability are important factors affecting the use of zinc. Markets and consumption for zinc die castings are given in Table 16.

Table 16. Markets for Zinc Die Castings, 1000 t^a

Year	Total consumption	Automotive components	Builders hardware	Electrical components	Domestic appliances	Industrial, agricultural, and commercial machinery	Miscellaneous industries	Sport goods and toys	Scientific, sound, and television equipment
1974	411.7	208.9	77.9	27.2	34.5	31.4	9.0	11.4	11.4
1976	377.8	187.4	79.5	33.5	27.4	21.5	9.8	8.9	9.6
1978	377.7	174.6	84.4	39.3	26.9	23.8	10.0	8.6	10.0
1980	284.7	94.0	76.0	37.8	22.8	25.4	11.6	9.3	7.8

^a Refs. 65,73.

Zinc Dust. U.S. production and imports are given in Table 17 (65), distribution of consumption in Table 18 (2). The slight decline since 1970 can be attributed to a reduced demand for chemical uses since the demand for zinc dust for coatings has increased.

Table 17. U.S. Production and Imports of Zinc Dust, 1000 t

Zinc dust	1970	1975	1978	1980	1995
production	46.0	38.2	38.5	42.6	
imports	8.5	5.2	9.0		11.7

Table 18. Distribution of Zinc-Dust Consumption, 1974–1977 Average

Use	1000 t yr	Percent
coatings		
industrial	17.1	35.2
automobile	10.0	20.5
marine	3.4	6.9
chemical uses	18.2	37.4
<i>Total</i>	<i>48.7</i>	<i>100.0</i>

Specifications and Grades

The chemical requirements for three slab-zinc grades made from ore or other material by distillation or electrolysis are specified by ASTM as given in Table 19 (74). Zinc produced by sweating or remelting of secondary zinc is not covered. Purity ranges from 98.0 to 99.99^o ¢; the main impurities are Pb, Fe, and Cd. The impurities strongly affect the physical and chemical properties. Over the years, the intermediate grades, brass special, and selected grades of slab zinc, with purities between high grade and Prime Western, have been deleted from the specification because use requirements have changed. Nevertheless, the use determines the quality standards (75).

Table 19. Compositions of the Commercial Grades of Zinc^a, %

Grade	Lead, maximum	Iron, maximum	Cadmium, maximum	Zinc, minimum by difference
special high grade ^b	0.003	0.003	0.003	99.990
high grade	0.03	0.02	0.02	99.90
Prime Western ^c	1.4	0.05	0.20	98.0

^a When specified for use in the manufacture of rolled zinc or brass, aluminum max 0.005%.

^b Tin max 0.001%.

^c Aluminum max 0.05%.

The ability of the zinc industry to produce special high grade by zinc distillation and electrolytic processes led to the development of modern zinc die-casting alloys which represent the second largest market for zinc. Because zinc die castings contain aluminum, this extra-pure grade must be extremely low in iron, lead, cadmium, and tin in order to avoid damaging intercrystalline corrosion. Such rigorous control of composition is essential for stable zinc die castings.

In galvanizing, the largest use of slab zinc, the iron or steel is dipped in molten zinc or electroplated. For hot-dip galvanizing, the lower grades are satisfactory. For some applications, such as wire, which must withstand bending without flaking of the coating, high grade or special high grade is employed. In electrogalvanizing, zinc is obtained either from zinc anodes or by direct solution from ore and subsequent purification of the electrolyte. High grade and special high grade are used for anodes in electrogalvanizing.

To meet the specification for Prime Western, zinc is debased with lead. This may increase the cost, when the price of lead is higher than that of zinc. Generally, lead, the principal impurity in Prime Western, aids the galvanizing process by increasing the fluidity of the bath and defining spangle boundaries. The solubility of lead in molten zinc at galvanizing temperatures (425–460°C) is ca 1–1.5% by weight. If the lead content is higher, the lead settles at the bottom of the galvanizing kettle, protecting it from attack by molten zinc and facilitating the removal of iron–zinc alloy dross.

All grades of zinc slab are used to some degree in brasses and bronzes. In many leaded brass-mill products, the lead originates from the slab zinc; the accompanying cadmium is usually acceptable.

All grades may be used in rolled zinc alloys, although special high grade and high grade are most commonly employed. When aluminum is added to superplastic zinc alloys (Zn-22Al), special high grade is required plus protective alloying elements to prevent intergranular corrosion caused by moist service conditions, as is the case with zinc die castings.

Analytical Methods

Zinc in ores at the concentrating mill is often determined polarographically and mill slurries are commonly monitored continuously for zinc by x-ray fluorescence to control addition of flotation reagent (76). Low zinc concentrations in solution are analyzed polarographically or by atomic absorption spectroscopy (AAS) with a detection limit as low as 0.01 µg mL (77). This method is used by most modern mills for determining other metals such as lead, cadmium, copper, cobalt, nickel, magnesium, and calcium. Silver and gold are fire-assayed. In concentrates where zinc is high and great precision is required, wet methods are often used, such as the titration with potassium ferrocyanide (78).

Zinc smelters use x-ray fluorescence spectrometry to analyze for zinc and many other metals in concentrates, calcines, residues, and trace elements precipitated from solution, such as arsenic, antimony, selenium, tellurium, and tin. X-ray analysis is also used for qualitative and semiquantitative analysis. Electrolytic smelters rely heavily on AAS and polarography for solutions, residues, and environmental samples.

Analysis of zinc solutions at the purification stage before electrolysis is critical and several metals present in low concentrations are monitored carefully. Methods vary from plant to plant but are highly specific and usually capable of detecting 0.1 ppm or less. Colorimetric process-control methods are used for cobalt, antimony, and germanium, turbidimetric methods for cadmium and copper. Alternatively, cadmium, cobalt, and copper are determined polarographically, arsenic and antimony by a modified Gutzeit test, and nickel with a dimethylglyoxime spot test.

Finished zinc and zinc alloys are usually analyzed for metals other than zinc by emission spectroscopy and the zinc determined by difference. ASTM method E 27 describes a technique using a dissolved sample and photographic detection. The internal standard is the zinc line at 267.0 nm. However, procedures using solid samples are generally preferred and photoelectric detection often replaces optical detection. Samples are cast and machined on the surface where the arc is struck. Up to 15 elements can be determined in a few minutes by modern automatic spectrometers. ASTM gives wet chemical methods for metals other than zinc (79).

The Committee on Medical and Biologic Effects of Environmental Pollutants of the National Research Council presents a well-referenced review of zinc sampling and analysis as they pertain to physiological media, water, air, foods, soils, and plants (80). Standard methods have not yet evolved in the clinical field but some are recommended. Most analyses today are by AAS which uses the 213.8-nm wavelength for zinc with reported sensitivities of 0.01–0.025 µg mL. Electrothermal atomization (graphite furnace) lowers the detection limit by up to three orders of magnitude (77). Spectrophotometric procedures are popular, commonly employing dithizone (diphenylthiocarbazone) and zincon to develop red (535 nm) and blue (620 nm), respectively. Those methods detect zinc concentrations as low as 1 µg mL (80). Polarography is still employed for water samples containing 0.01–0.1 mg Zn L, but is being replaced by AAS.

Health and Safety Factors; Environmental Aspects

Within recent years there have been a number of review articles regarding the release, fate, and transport of zinc in our environment. The publications have provided the insight not only to the environmental hazards of zinc but also to the biological importance of this essential element as well (81–85).

The level of natural versus man-made emissions to the environment are of a similar magnitude. Soil erosion is the major contributor of natural emissions with zinc mining, zinc production facilities, iron and steel production, corrosion of galvanized structures, coal and fuel combustion, waste disposal and incineration, and the use of zinc fertilizers and pesticides being the principal anthropogenic contributors.

The concentrations of zinc in air, water, and soil are highly variable depending on geological considerations and the influences of point sources in the area. The natural background concentrations of total zinc range from <0.1–50 µg L in fresh water, 0.002–0.1 µg L in seawater, 10–300 mg kg dry weight in soil, and up to 300 ng m³ in air.

Zinc is an essential element and thus can exist in both the deficient or toxic state in plants and animals. Deficiency occurs in some plants when the tissue content drops below 20 ppm. The normal range is 25–150 ppm. The toxic effect of zinc has been seen at 400 ppm, although plants vary widely in their tolerance.

Direct application of zinc to soil to optimize growing conditions is accomplished using commercial-type fertilizers or through the application of municipal sewage sludge. The latter is closely regulated in order to prevent metals, including zinc, from accumulating to levels that would become phytotoxic or result in an undue exposure to humans through the ingestion of food or soil.

Like plants, marine and fresh water organisms vary significantly in their response to zinc. The range of concentration among species is relatively narrow, except for oysters, which tend to accumulate zinc to very high levels. The zinc content of most seafoods ranges from 3–30 ppm. Oysters contain from 100–2000 ppm.

Zinc is also an essential food element in the human diet. Too little zinc in the diet can lead to poor health, reproductive problems and a lowered ability to resist disease. Taking too much zinc into the body through food, water or dietary supplements can also affect health. The levels of zinc that produce adverse effects are higher than the Recommended Daily Allowances, which are 15 mg day for men, 12 mg day for women, 10 mg day for children and 5 mg day for infants.

Hazards of Production. In most zinc mines, zinc is present as the sulfide and coexists with other minerals, especially lead, copper, and cadmium. Therefore, the escape of zinc from mines and mills is accompanied by these other often more toxic materials. Mining and concentrating, usually by flotations, does not present any unusual hazards to personnel. Atmospheric pollution is of little consequence at mine sites, but considerable effort is required to flocculate and settle fine ore particles, which would find their way into receiving waters.

Drainage from active mining areas is considerably less than from inactive mines because of the disposal methods currently employed. Prior to discharge, liquid effluents are limed and settled to precipitate metals as hydroxides. Flocculants are used to reduce the total suspended solids and, in some instances, filtration of thickener overflow is practiced.

Particulate emissions are controlled mainly through venting, baghouses and water scrubbers. Atmospheric zinc loss is estimated at 100 g t or zinc mines, mostly from handling dry ore and concentrate and wind erosion of tailing piles. Sulfur dioxide emissions have been reduced by installing double absorption acid plants and improved containment of dilute gases.

The most significant occupational exposures to zinc would occur during the smelting and refining of zinc ore. The standards for occupational exposure have been established at a level to prevent the onset of metal fume fever. This temporary condition is caused by excessive exposure to freshly formed fumes of zinc oxide and results in flulike symptoms of fever, chills, headache, muscle pain, nausea and vomiting.

Uses

Metallic Coatings. Zinc coatings on iron and steel for corrosion protection represented the largest market for zinc in the 1970s (see Table 15). Methods in general include hot-dip galvanizing, continuous-line galvanizing, electrogalvanizing, zinc plating, zinc spraying, and painting with zinc-bearing paints (86).

Die-Casting Alloys. The principal application of zinc as a structural material is in alloys for pressure die casting (see Table 16). In 1992, markets were equally divided between automotive, builders' hardware, and all other uses. The automotive industry casting uses include handles and locks, mechanical components, electrical components, body hardware and trim, lamp and lighting fittings, instruments, and other components. The zinc die castings per average car, reflecting the need for lighter weight and thin-walled castings, dropped from ca 22.6 kg per car in 1974 and 1975 to 13.6 kg in 1978, 12.8 kg on the 1979 models. Builder's hardware uses include covered door and window hardware, locks and keys, furniture and cabinet hardware, hand tools and cutlery, bathroom and plumbing fittings, general hardware, marine hardware, and luggage hardware.

The three zinc alloys, designated as Nos. 3, 5, and 7, are used for the hot-chamber die-casting process. Their compositions are given in Table 20 along with a cold-chamber zinc die-casting composition, No. 16 alloy (87–89). The No. 16 alloy was developed by the International Lead Zinc Research Organization (ILZRO) and has superior creep resistance at both ambient and elevated temperatures (90). Alloys ZA-87, ZA-12, and ZA-27 described later can also be die cast. ZA-8 is growing in use as a hot-chamber die-casting alloy, whereas ZA-12 and ZA-27 are cold-chamber die cast.

Table 20. Compositions of Zinc Die-Casting Alloys, wt %^a

Element	No. 3 ASTM AG40A		No. 5 ASTM AC41A		No. 7		No. 16	
	Ingot ^b	Die casting ^{c,d}	Ingot ^b	Die casting ^{c,d}	Ingot ^b	Die casting ^c	Ingot	Die casting
copper	0.10 max	0.25 max ^e	0.75–1.25	0.75–1.25	0.10 max	0.25 max ^e		1.0–1.5
aluminum	3.9–4.3	3.5–4.3	3.9–4.3	3.5–4.3	3.9–4.3	3.5–4.3		0.01–0.04
magnesium	0.03–0.06	0.03–0.08 ^f	0.03–0.06	0.03–0.08 ^f	0.010–0.020	0.005–0.020		0.02
iron, max	0.075	0.100	0.075	0.100	0.075	0.100		0.04
lead, max	0.005	0.007	0.005	0.007	0.0020	0.0030		0.005
cadmium, max	0.004	0.005	0.004	0.005	0.0020	0.0020		0.004
tin, max	0.002	0.005	0.002	0.005	0.0010	0.0010		0.003
nickel					0.005–0.020	0.005–0.020		
titanium								0.15–0.25
chromium								0.10–0.20

^a Zinc is the remainder.
^b ASTM B 240.
^c May contain nickel, chromium, silicon, and manganese in amounts of 0.02, 0.02, 0.035, and 0.5° ‰, respectively. No harmful effects have ever been noted due to the presence of these elements in these concentrations; therefore, analyses are not required for these elements.
^d ASTM B 86.
^e For most commercial applications, a copper content in the range of 0.25 to 0.75° ‰ does not adversely affect the serviceability of die castings and should not serve as a basis for rejection; ASTM B 240-79 (93).
^f Magnesium may be as low as 0.015° ‰ provided that the lead, cadmium, and tin do not exceed 0.003, 0.003, and 0.001° ‰, respectively.

The zinc alloys 3, 5, and 7 are somewhat similar in properties and can generally be used interchangeably. The No. 3 alloy has excellent retention of impact strength and dimensions; No. 5 has greater hardness, tensile strength, and more resistance to creep, but suffers some loss of impact strength when used continuously at elevated temperatures; No. 7, the most recent addition to the series, has properties similar to No. 3, but is slightly softer and more ductile. Because of its lower magnesium content, it exhibits better fluidity and castability than either 3 or 5 (91).

Aluminum is the principal alloying addition varying from 3.5 to 4.3° ‰. The Zn–Al phase diagram is shown in Figure 4 (92). Most die-cast and foundry alloys are Zn–Al compositions. Aluminum improves die-cast strength, reduces grain size, and decreases the solution rate of iron and steel parts to the extent that a submerged plunger-type die-casting machine can be used. The spread between 3.5 and 4.3° ‰ allows sufficient latitude for production without decreasing optimum properties. An aluminum content below 3.5° ‰ promotes hot shortness, impairs the surface finish and lowers the mechanical properties. On the high aluminum side, a loss of impact strength begins at 4.5° ‰ aluminum, and at 5° ‰ the alloy is brittle. Aluminum in zinc, in the presence of excessive amounts of certain impurities (eg, Pb, Sn), can result in intergranular corrosion in die castings exposed to moist, warm atmospheres, causing them to lose strength, crack, and sometimes even disintegrate. Specifications place limits on these harmful impurities to avoid this problem, and special high grade zinc is employed for die-casting alloys (94,95).

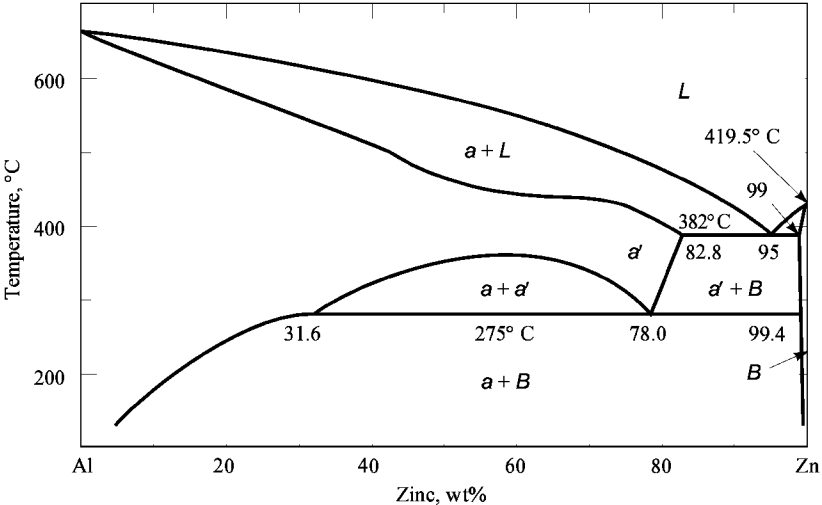


Fig. 4. Aluminum-zinc phase diagram.

Copper increases tensile strength and hardness and offers some protection against elements that promote intergranular corrosion. However, copper reduces impact strength and dimensional stability owing to aging and is therefore kept at 1.25% max.

Magnesium prevents intergranular corrosion by counteracting the harmful effect of small quantities of impurities such as tin and lead (96). The minimum values for magnesium are necessary for the level of impurities allowed in the casting specification. The maximum limit for magnesium in the alloy specifications is set arbitrarily to allow some spread for commercial production. Operating close to the minimum values lessens the hot shortness and improves castability. The No. 7 alloy allows the lowest minimum magnesium content in the presence of nickel which neutralizes elements that promote intergranular corrosion. Magnesium also increases the strength and the hardness.

Maximum limits are placed on iron, lead, cadmium, and tin. The maximum limits for lead and tin are set at amounts that do not promote subsurface network corrosion at the minimum levels of magnesium allowed in the die castings. At concentrations of ca 0.1% and higher, cadmium is detrimental to mechanical properties. Cadmium adversely affects hot shortness and castability, but the amount in special high grade zinc is too low to produce a noticeable effect. For the Nos. 3 and 5 alloys, the maximum limit is 0.004%, but for the No. 7 alloy, it is 0.002%, and special high grade zinc must be selected to meet this requirement in the preparation of this alloy. Iron has no detrimental effects on the permanence or properties, but an excess of iron may affect finishing operations such as machining and buffing. Zinc castings should not contain much over 0.02% iron. The ASTM limit of 0.100% iron is an arbitrary value which may be on the high side. The ASTM specification allows small quantities of nickel, chromium, silicon, and manganese which may come from aluminum used in alloying or from remelting of electroplated scrap (94).

The melting of zinc alloys for die casting entails (95) the melting down of virgin ingot; remelting of scrap from the foundry and trimming operations and reconstitution of this melt by means of suitable additions; and the holding of quantities of molten metal at a closely controlled temperature adjacent to the die-casting machine. The furnaces are generally of the pot type, although immersion-tube and induction furnaces are also used. Heat is supplied by gas, oil, or electrical resistance. The capacity of melting furnaces is between 450 and 9000 kg; total capacity is usually five to seven times the amount of metal required per hour. The pots are made from gray or ductile cast iron. Ladles are cast iron or pressed steel. The temperature of zinc is kept at 475–500°C in the melting and alloying pots, whereas holding furnace temperatures are between 390 and 425°C, depending on the composition of the alloy being cast and part size.

Die castings are produced under pressure in permanent metal molds (97–99). Die casting with zinc-base alloys is one of the most efficient and versatile production methods which can be used for the manufacture of accurate, complex metal components. The machine used is of the so-called hot-chamber or submerged-plunger type because the alloys melt at low temperature and do not attack the injection-pump material (gray or ductile cast iron). The die-casting die or mold consists of a cover die and an ejector die which meet at the parting line. Machines range in size (clamping force) from 25 to 2,500 tons.

Injection pressures for hot-chamber machines are between 6.9 and 20.7 MPa (1000–3000 psi). The velocity with which the metal enters the die is controlled by the pressure and die design. The optimum may differ from each casting and in the past it has been determined primarily by trial and error. In recent years, techniques have been developed for the use of instrumentation to calibrate the shot system and die design programs are available for optimizing gate and runner systems in dies (100,101). The P–Q² technique (relationship between pressure and flow), developed by the Commonwealth Scientific and Industrial Research Organization, is utilized to characterize the rate with which the die-casting machine-shot system delivers zinc to the die. Using this information, metal flow during die filling can be controlled for castings of various sizes (102). Computer programs are available to aid the die caster in applying these rules and guidelines (103).

For casting zinc alloys, die temperatures are generally between 150 and 250°C. The lower temperatures are employed for heavy castings, higher temperatures for thin-wall castings. Hot-chamber machines can be operated rapidly with a high degree of automatic control, generally at rates of 50–500 shots per hour. Special machines greatly exceed these rates with 2000–5000 shots per hour up to 18,000 per hour for a zipper-casting machine. Castings may weigh from a few grams to over 22.5 kg.

Zinc alloys can be cast in single-cavity, multiple-cavity, combination, or unit dies. Combination or family dies consist of a series of cavities in one die for casting two or more parts. Unit dies are separate, small dies, usually with a single cavity, which are inserted in a single master-holding die and operate several at a time in large machines. Because temperatures for die casting are relatively low, 150–250°C, hot-worked tool steels are not generally required. For short-run dies, low alloy steels, like 4140 steel, are suitable. However, for very long runs, particularly when high dimensional accuracy is required, hot-worked tool steels, such as H11, H12, and H13, give the longest die life. Ejector pins of nitrided H11 tool steel or of 7140 alloy steel are available as stock items for insertion in the dies. Hardenable grades of stainless steel, such as type 440B, are often used for cores. Hot-worked tool steels such as H11, H12, and H13 can be used for both cores and slides. Lubricating the slides and cores with molybdenum disulfide or colloidal graphite in oil helps to ensure smooth action and to minimize wear.

The life of a casting die depends on the temperature of the metal being cast, thermal gradients within the die, and frequency of exposure to high temperature. The die life for casting zinc alloys is generally much longer than that for casting aluminum, magnesium, or copper alloys; it is not unusual for dies to last for one million (10⁶) shots or more.

Zinc die castings may be machined, bent, swaged, or coined for finishing or for slight changes or shape (94). They can be joined by riveting, spinning, welding, and soldering in assembly operations. Corrosion resistance against atmospheric and seawater conditions is good, particularly if first treated with chromate using a weak chromic acid solution to form a passive film on the surface. Anodized coatings protect zinc against severe corrosive environments and improve wear resistance. Paint, lacquer, or enamel finishes can be applied for decorative purposes. A large number of parts are chromium plated. For such treatment, the castings are usually buffed to remove traces of parting lines and flow marks, and then plated with copper, nickel, or chromium. Other metals can also be electroplated on zinc.

During the 1970s, casting techniques for the production of thin-wall zinc die castings were developed by the ILZRO to allow the faster production of components containing less zinc at lower cost (100). Because of the resulting weight saving per part, zinc became cost- and weight-competitive with aluminum and plastic (104,105).

Castings with exceptionally thin wall sections are produced by faster cooling rates, controlled metal injection, and a meticulous standard of die construction which is necessarily associated with automatic machine operation. Since die casting is a heat-transfer process, any reduction in the mass of metal to be chilled in each cycle allows faster operation. For the user, metal saving has not only a financial advantage, but gives a component with better strength-to-weight ratio since thin sections have a fine-grained structure with superior properties. Thus, castings up to ca 3 kg can be made with walls ca 0.75 mm thick (106,107).

Purging the die cavity with a reactive gas before metal injection eliminates gas porosity which is found in die castings because of trapped air. In the pore-free process, the air in the die cavity is displaced with oxygen or other reactive gas. When the die is charged with the molten zinc alloy, the zinc reacts with the gas to form tiny, solid particles that are dispersed throughout the pore-free casting. The absence of porosity in the castings results in improved mechanical properties. This process is still in the developmental stage for zinc die castings (108).

Foundry Alloys. The tonnage of slab zinc used in foundry applications is small compared with that used for die casting, 5334 t in 1979 and 1200 MT in 1996.

High Strength Alloys. Before 1967, the only zinc gravity-casting alloys of note were the slush and forming-die alloys. In that year, the first high strength zinc gravity-casting alloy, No. 12, was introduced with 11% Al. This alloy was originally marketed as a prototype alloy to be used in sand or permanent-mold casting of parts that would later be made as die castings, since many of its mechanical properties approximate those of die-cast alloy No. 3. However, much wider application of this alloy in the foundry industry has developed (109). Subsequently, two additional alloys were developed containing 8% aluminum (No. 8) and 27% aluminum (No. 27) (110); compositions are given in Table 21 (89,111,112). These alloys can also be pressure die cast.

Table 21. Compositions of High Strength Zinc Foundry Alloys, wt %^a

Element	No. 8	No. 12 ^b	No. 27
aluminum	8.0–8.8	10.5–11.5	25–28
copper	0.8–1.3	0.5–1.25	2.0–2.5
magnesium	0.015–0.03	0.015–0.03	0.01–0.02
iron	0.10	0.075	0.10
lead	0.004	0.004	0.004
cadmium	0.003	0.003	0.003
tin	0.002	0.002	0.002
zinc	balance	balance	balance

^a Single units indicate maximum amounts permitted.
^b Covered by ASTM B 669-80 (109). A revision of this specification is under consideration to include Nos. 8 and 27.

Physical and mechanical properties are given in Table 22 (109,110,112–114). The densities reflect the effect of aluminum: the Zn–27° Al alloy is ca 30° lighter than zinc, 17° lighter than the Zn–11° Al alloy, and 21° lighter than the Zn–8° Al alloy. All three alloys possess a good combination of mechanical properties. The sand-cast Zn–27° Al alloy is superior to the other two alloys with a tensile strength close to 50° higher. Ductility of the Zn–27° Al alloy is slightly higher; however, if ductility is required, the heat-treated condition should be specified. The Zn–27° Al alloy is also the most creep-resistant (90,113). Mechanical properties of these alloys are equal or superior to those of many brass, bronze, aluminum, and iron alloys.

Table 22. Properties of High Strength Zinc Foundry Alloys^a

Property	Alloy No. 8, Permanent-Mold Cast	Alloy No. 12		Alloy No. 27	
		Sand Cast	Permanent-Mold Cast	Sand Cast ^b	Sand Cast, H.T. ^c
physical					
density, g cm ³	6.37	6.03	6.03	5.01	5.01
melting or solidifi- cation range, °C	375–404	377–432	377–432	375–487	375–487
mechanical					
tensile strength, MPa ^d	221–225	276–310	310–345	400–441	310–324
yield strength, 0.2° MPa ^d	207	207	214	365	255
elongation, °	1–2	1–3	4–7	3–6	8–11
Brinell hardness ^e	85–90	92–96	105–125	110–120	90–100

^a Refs. 109, 110, 112–114.
^b Primary purpose.
^c Heat treated.
^d To convert MPa to psi, multiply by 145.
^e 500 kg for 30 s.

The Zn–Al system permits manipulation of the mechanical properties by suitable heat treatment. The aluminum-rich alpha phase is especially suitable for solution hardening since it can be supersaturated by as much as 30 wt ° zinc. Furthermore, both alpha and beta phases can be strengthened by precipitation because of decreasing solute solubility with decreasing temperature.

Alloys Nos. 8, 12, and 27 should not be prepared and stored in cast-iron crucibles since excessive iron pickup decreases fluidity. Many types of refractory crucibles can be used for molten zinc–aluminum alloys. Silicon carbide crucibles are generally recommended. However, unlike other alloys, zinc alloys can be cast in many types of molds, including sand, plaster, silicone rubber, graphite, and cast iron (109,114), as well as bronze, aluminum, and beryllium–copper molds. Silicone rubber and graphite molds offer some special advantages. Parts with reverse taper can be cast in silicone rubber molds because of their flexibility. Graphite is easy to machine, does not distort or warp, and promotes rapid solidification because of high thermal conductivity; over 20,000 zinc alloy castings have been made with the same graphite mold (115). Castings made in graphite molds show superior surface finish and mechanical properties. Alloy No. 12 can be sand- and permanent-mold cast and graphite molds are particularly successful. Alloy No. 8 has excellent finishing characteristics and presents a low cost replacement for brass and bronze castings.

The gravity-casting alloys are gaining widespread acceptance for industrial structural parts; business machines, office equipment, builder's hardware, and marine applications (114–116). They have replaced cast iron, copper-based alloys, and aluminum alloys in many applications, eg, Alloy No. 12 is used in bearing applications, where bushings have been tested in mine elevators, drill motors, 50-ton dump trucks, and drill-track rollers. The alloy bushings showed less wear than bronze when both were used in tandem on the same equipment. It is expected that zinc-based bearings and bushings will find wider acceptance as a substitute for bronze because of their lower cost and superior mechanical properties, particularly for low speed, high-load-type service (112,117).

Because zinc-based alloys have low melting points, energy savings in the melting operation are substantial and the foundry operation is essentially free of fume. With the current trend of increasing energy costs and pollution control, cost benefits can be considerable (114).

Slush Alloys. Slush casting is limited generally to the production of hollow casting (118), eg, lamp bases, lighting fixtures, and casket hardware. In slush casting, a molten metal is poured into a split metal mold, generally made of bronze, until the mold is filled; the mold is immediately inverted and the liquid metal is allowed to run out, leaving a thin-shell casting behind, whose shape is a replication of the cavity walls. The thickness of the wall of the casting depends on the time interval between the filling and the inverting of the mold, as well as on the chemical and physical properties of the alloy and the temperature and composition of the mold. Skilled operators produce good castings with remarkable uniformity in weight.

Slush-casting alloys must be fairly low melting and freeze over a temperature range, that is, the liquidus and solidus temperatures must be significantly different to provide a slushy range.

Zinc slush-casting alloy compositions are based on the Zn–Al system. The two commonly used alloys have nominal aluminum contents of 4.75° Al and 5.5° Al, which fall on either side of the zinc–aluminum eutectic at 5° Al with an invariant melting point of 382°C (see Fig. 4). Special high grade zinc is employed to control impurity levels in order to avoid corrosion problems. The lower aluminum alloys with 0.25° Cu produces castings with improved mechanical properties and longer service life, because copper affords protection against corrosion. Slush casting with copper-containing alloy, however, is more difficult and yields castings with thicker walls than the straight binary Zn–5.5° Al alloy.

The cast zinc–aluminum–copper slush alloy, after aging for ten years indoors, shows a tensile strength of 238 MPa (34,500 psi) with Charpy impact strengths of 1.4–4.1 J (1–3 ft-lbf) (87).

Forming-Die Alloys. The tonnage of slab zinc used in this application is small. The use of zinc alloy dies started in the aircraft industry during World War II (119). Zinc-based alloys cast in sand and plaster molds continue to be used for short-run dies for steel and aluminum stampings in the automotive and aircraft industries (120). Considerable cost savings are realized with these low melting zinc-based alloys which are easy to polish, machine, weld, and remelt.

The common composition is 4° Al and 3° Cu, with or without a small amount of magnesium. Other elements, such as Ni or Ti, increase die life (121). The base alloy has tensile strengths of 210–280 MPa (30,450–40,600 psi) and Charpy impact strengths of ca 20–27 J (15–20 ft-lbf) (90).

Rolled Zinc. Rolled-zinc products constitute ca 2–3° of slab zinc consumed in the United States over the past ten years. The U.S. production has declined by ca 50° since 1965 (65) and in 1996 only two companies rolling zinc products were left.

The ASTM B 69-66 (1979) gives typical compositions of rolled zinc for informational purposes and general guidance and not for specification purposes (122). The maximum solid solubilities of other metals in zinc are always small, 0.5–1° for manganese and aluminum; ca 2–3° for cadmium, palladium, and copper; ca 8° for silver; and 10–15° for gold; solubilities at room temperature are lower. Only fractional percentages, or minute amounts, of other metals enter the structure of the zinc crystal, and only copper, cadmium, and aluminum are added to modern zinc rolling alloys. The commercial grades of rolled zinc contain the natural impurities lead, cadmium, and iron. Rolled zinc compositions developed to meet various applications are given in Table 23 (89,123). Except for the superplastic Zn–Al, the total alloying content in the other alloys is <2%. Limits are placed on impurities like cadmium, lead, iron, tin, arsenic, bismuth, and indium which cause hot shortness or edge cracking. Lead, tin, cadmium, and bismuth must be held at very low concentrations in aluminum containing alloys because they promote intercrystalline corrosion, which takes the form of exfoliation of the surface and general embrittlement; therefore, special high grade zinc is used in the preparation of these alloys.

Table 23. Rolled Zinc Alloys

Alloy	Pb	Cd	Fe	Cu	Ti	Mg	Al	Zn
zinc, pure	0.003–0.10	0.003–0.007	0.0014–0.012	0.001 max				99.88–99.99°
	max	max	max					
Zn–Pb–Cd–Fe	0.04–0.40	0.035–0.35	0.006–0.030	0.002–0.005				99.2°
			max	max				
Zn–Cu	0.02–0.10	0.007–0.08	0.012 max	0.12–0.90				98.9°
	max							
Zn–Cu–Mg	0.10 max	0.04 max	0.012 max	0.65–0.85		0.006–0.016		97.98°
Zn–Cu–Ti	0.003–0.30	0.007–0.02	0.002–0.012	0.50–0.90	0.08–0.16			98.61°
		max	max					
Zn–Al–Mg	0.002–0.003	0.002–0.003	0.002–0.003	0.002 max		0.03–0.07	0.08–0.11	99.81°
	max	max	max					
Zn–HiAl, super-plastic	0.003 max	0.003 max	0.003 max	0–0.6		0–0.03	20–24	75.36°

The zinc is normally melted in a gas, oil, or coal-fired reverberatory furnace with a capacity up to 100 tons or in a low frequency induction furnace with a capacity of a few tons. The more highly alloyed compositions are more effectively melted and mixed in low frequency induction furnaces. The furnace must be refractory-lined to eliminate iron pickup by the molten metal. The metal temperature is maintained below 500°C to minimize loss by oxidation. A ladle is used to transfer the metal for casting into molds; the pouring temperature is usually ca 440°C. Zinc scrap is not generally suitable for remelting because it may contain undesirable impurities.

Although more and more zinc sheet and strip are produced in continuous mills, some is still produced by rolling slabs cast in open or closed book-type molds made of cast iron (124–127). The casting temperatures are between 440 and 510°C, mold temperatures between 80 and 120°C. The contact surfaces of the mold must be smooth and clean to allow unrestricted shrinkage of the cast slab. Mold lubricant is not necessary, but if used should be held to a minimum. Slabs cast in open molds must be skimmed immediately to remove surface oxide. Rolling slabs are cast 1.87–10 cm thick.

Zinc rolling slabs have been cast successfully by semicontinuous direct-chill casting methods. This is the preferred method for superplastic zinc alloys which, because of their large freezing range, display unacceptable surface shrinkage when cast in open molds.

The open-mold cast slabs are preheated to 150–260°C for initial rolling in rough or breakdown mill. Composition of slab influences the temperature employed. Special high grade or high grade is used to avoid the harmful effect of impurities. Tin should be absent altogether, because its presence at concentrations of 0.004° can cause the zinc to crumble at hot working temperatures. Slabs for strip are reduced by longitudinal rolling to a thickness convenient for coiling, generally <2.5 mm. For sheet, slabs are rolled laterally until the desired width is obtained. Then the sheets are turned 90° and rolled to a thickness two to four times the final gauge. By cross rolling, the slab is worked in two directions, reducing to some degree the effects of anisotropy.

Strip is produced as wide as 2 m and in thicknesses as low as 0.1 mm in regular mills. Foil in thicknesses of 0.025 mm or less is produced in special mills. To obtain strip with a bright surface, high ductility, and low hardness, finish rolling is performed hot at 120–150°C.

Zinc sheet is produced by the pack-rolling process, in which 2–40 rough-rolled sheets are stacked together in packs and rolled simultaneously. Packs must be split frequently, interchanging inner and outer sheets to equalize temperatures and reductions. With care, good quality sheets can be produced by this technique but considerable variations in properties can occur. If a bright ductile product is desired, rolls are held at 120–150°C; the reduction on the last pass is 20–40°.

Lubrication of sheet and strip is necessary for all operations. Although for special operations vegetable and mineral oils may be employed, a mixture of paraffin and tallow oil is normally preferred in rough rolling. Requirements for finish-roll lubricant are more strict because of staining caused by breakdown of the oil or reaction with the zinc. Strip zinc is usually finish-rolled with cotton seed or mineral oil.

In recent years, much attention has been given by the zinc industry to the development of methods for casting and rolling the metal continuously (128–130). The Hazelett casting machine employed consists essentially of two continuous steel belts, supported and driven through a system of pulleys (131). These form the walls of the mold, which is closed at the sides by sets of small steel blocks linked together by cables forming two endless chains sandwiched between the belts. Molten metal is fed from a tundish or funnel into the cavity between the belts, which are cooled by water sprays. The belts are rotated at a controlled speed and, as the metal solidifies, the slab is drawn downward, thus producing a continuous cast ingot; its thickness can be varied from 10 to 75 mm by adjusting the belt separation. Owing to the continuous nature of the casting operation, the grain size of the solid slab is much finer than when static molds are used, and subsequent reduction in the rolling mill is therefore easier, giving a more uniform product.

After leaving the casting machine, the slab is cooled by water sprays to 180–240°C and fed into the mill, which is generally of the four-high roll type, where a 60° reduction is taken in one pass. The strip is then coiled and, when the casting run has been completed, is fed back through the mill and rolled at a temperature of 80–90°C, with a reduction up to 50°, to give the required final thickness, finish, and properties.

Zinc sheet and strip can be easily fabricated by drawing, spinning, bending, stamping, coining, embossing, blanking, and impact extrusion. Rolled-zinc products in the form of strip, sheet, wire, and rod have many and varied commercial applications. Strip is formed into dry-cell battery cans, mason jar covers, organ pipes, grommets, eyelets, and many other objects, some of which are subsequently brass or chromium plated (jewelry, medallions, bathroom accessories, etc) (132). The zinc–carbon dry-cell application accounts for about one half the rolled-zinc consumption in the United States (see BATTERIES). Sheet zinc is used in photoengraving and also in the construction of roofing and other architectural uses. Special high grade zinc with a

maximum iron content of 0.0014% in the form of plate and rods gives cathodic protection to steel in marine and pipeline applications.

The low creep resistance of unalloyed zinc sheet and strip is one of its most serious defects and restricts its application. Strip made with Zn–0.6 Cu–0.10 Ti composition has superior creep resistance and is used in structural applications where high, continuously applied stresses are encountered in service.

A new large tonnage application for strip zinc has developed from the decision in 1982 of the U.S. Mint to replace the solid copper (95% Cu–5% Zn) penny with a less expensive copper-plated penny made from zinc strip (Zn–0.8% Cu alloy) (133,134). Zinc blanks stamped from the Hazelett-cast rolled zinc are barrel-plated with copper prior to coining to produce the finished penny. In 1992, about 22,000 Mt of special high grade zinc were used to make 9.1 billion pennies.

Another commercial development of the 1970s is the application of superplasticity which is exhibited by a number of zinc alloys (135–138). Under the right conditions, the material becomes exceptionally soft and ductile and, under low stresses, extensions exceeding 1000% can be obtained without fracture. The grain size must be extremely small (about 1 micrometer) and stable. This grain size is less than one tenth that of common metals in the wrought condition.

Extremely fine-grained two-phase structures can be developed in zinc–aluminum eutectoid alloys (ca 22% aluminum with and without minor alloying additions) by rapid quenching from above the eutectoid temperature of 275°C or by rolling to appreciable degrees of deformation at somewhat lower temperatures, or both. If sheet with this fine-grained structure is heated to just below the transformation temperature of 245–270°C, optimum superplastic behavior is obtained. Because of the ductility and low strength in this condition, the material is similar to thermoplastic polymers and, similarly, can be molded into complex shapes by various processes. When cooled to room temperature, its strength and hardness are many times greater than those of any thermoplastic. Tensile strengths of the superplastic zinc alloys at room temperature range from 190 MPa (28,000 psi) for the binary alloy to 380 MPa (55,000 psi) with the higher alloy compositions. A simple postforming heat treatment can increase strength to 350–440 MPa (51,000–64,000 psi). These alloys can be molded easily like thermoplastics at moderately elevated temperatures, but at room temperature have normal metallic properties, including strength, stiffness, conductivity, and stability, which no plastic can equal. The tooling cost is generally considerably lower than for conventional pressing, which is particularly beneficial for limited production runs (10,000 parts or less).

Parts made from these materials are formed from sheet metal, or solid, three-dimensional metal preforms (139). Complex housings, bezels, and equipment covers are typical of the former. Applications involving solid shapes include gears, pulleys, flywheels, and similar parts.

Zinc Dust and Powder. Zinc dust and powder are particulate forms of zinc which have many interesting and useful applications (2). The terms dust and powder have been used more or less indiscriminately to designate particulate zinc materials. Here, the term zinc dust designates material produced by condensation of zinc vapor, whereas zinc powder indicates the product obtained by atomizing molten zinc. Zinc dust is manufactured by vaporizing zinc from a suitable source material. The vapor is condensed under conditions which permit control over purity and fineness (140–142). The zinc vapor emanating from a boiler is led into a condensing chamber where it is diluted with inert gases and quickly chilled. The frozen droplets or dust are collected at the bottom of the condenser, screened, and packaged. The purity of the zinc dust depends on the source material or charge to the zinc boiler.

In the atomizing process, a stream of molten zinc is broken into tiny droplets by the force of a pressurized fluid impinging on the stream. The fluid can be any convenient material, although air is normally used. The atomized drops cool and solidify rapidly in a collection chamber. The powder is screened to specified sizes. Particulate zinc is also produced by other methods such as electrolytic deposition and spinning-cup techniques, but these are not of commercial importance.

Zinc dust is smaller in particle size and spherical in shape, whereas zinc powder is coarser in size and irregular in shape. The particle size of zinc dust, important in some applications, is controlled by adjusting the rate of condensation. Rapid cooling produces fine dust, slower condensation coarse dust. In the case of zinc powder, changes in the atomization parameters can be employed to change particle size to some degree. The particle size distributions for commercial zinc powders range from 44 to 841 μm (325–20 mesh). The purity of zinc powders is 98–99.6%.

ASTM recognizes two types of zinc dust in specification ASTM D 520-51 (reapproved 1976) (143), which includes permissible impurity concentrations. The metallic content of most commercial grades is 95–97%. The zinc oxide content is between 3 and 5%; finer dusts contain higher concentrations because of high surface areas. Zinc dusts are manufactured in various size ranges, and a typical commercial dust has an average particle diameter between 4 and 8 μm. Usually, dusts are screened to be essentially free of particles coarser than 75 μm (200 mesh).

The principal use of zinc dust is in paint coatings (see Table 18). Zinc-dust paints can be classified into those containing zinc oxide as a substantial portion of the pigmentation and those in which zinc dust represents most of the pigment (see PAINT). In the latter, the zinc-dust concentration in the dry paint film is high enough to provide galvanic protection to an iron or steel substrate similar to a galvanized coating (144,145). These paints are classified as zinc-rich paints, and represent one of the fastest growing new markets for zinc in recent years. Such paints are used to supplement galvanized steel for automotive underbody protection against corrosion (146), bridges and other structures, ship hulls, and buildings (2,147,148). Zinc-rich paint is recommended whenever conventional galvanizing is not suitable.

Zinc dust is used in the sherardizing process where work pieces are tumbled with zinc dust in rotating steel drums which are heated electrically or by gas to 370–420°C (149). The steel parts are uniformly coated with zinc. In the chemical and metallurgical industries, zinc dust is used as a reducing agent, in the manufacture of hydrosulfite compounds for the textile and paper industries, and to enhance the physical properties of plastics and lubricants (2).

By far, the largest application of zinc powder is for solution purification in electrolytic zinc plants. This application consumed an estimated 17,700 t of powder in 1980. Zinc powder is also used in primary batteries, frictional materials, spray metallizing, mechanical plating, and chemical formulations.

By-Product Metals

Ores that are exploited primarily for zinc invariably contain one or more other valuable metals. Various ores contain different combinations of such other metals (see Table 24). For instance, eastern ores commonly contain only cadmium as a significant recoverable value.

Table 24. Other Metals in North American Zinc Concentrates

Metal	Percent ^a	Process step	
		Electrolytic	Pyrometallurgical
lead	0.2–6.0	leach residue	sinter fume, refining column, ISF lead bullion and copper dross
copper	0.1–2.8	leach residue, purification residue	retort residue, ISF lead bullion and copper dross
cadmium	0.1–0.8	purification residue	sinter fume, refining column
silver	27–1900 ^b	leach residue	retort residue, ISF lead bullion and copper dross
gold	0.3–12 ^b	leach residue	retort residue, ISF lead bullion and copper dross
mercury	0.001–0.03	roaster off-gas	roaster off-gas
germanium	0.004–0.04	leach residue	sinter fume
gallium	0.005–0.02	leach residue	retort residue
indium	0.01–0.03	leach residue	sinter fume, refining column
thallium	trace	leach residue	sinter fume

^a Unless otherwise indicated.

^b g t.

Cadmium and mercury are usually recovered in separate processes at the zinc plant. The others are shipped as enriched residues to plants that specialize in their recovery.

BIBLIOGRAPHY

"Zinc and Zinc Alloys" in *ECT* 1st ed., Vol. 15, pp. 224–275, by A. Paul Thompson and H. W. Schultz, The Eagle-Picher Co.; "Zinc and Zinc Alloys" in *ECT* 2nd ed., Vol. 22, pp. 555–603, by A. W. Schlechten, Colorado School of Mines, and A. P. Thompson, Eagle-Picher Industries, Inc.; "Zinc and Zinc Alloys" in *ECT* 3rd ed., Vol. 24, pp. 807–851, by Thomas B. Lloyd, Gulf and Western National Resources Group, and Walter Showak, The New Jersey Zinc Company, Inc.

1. H. O. Hoffman, *Metallurgy of Zinc and Cadmium*, McGraw-Hill Book Co., New York, 1922.
2. B. C. Hafford, W. E. Pepper, and T. B. Lloyd, *Zinc Dust and Zinc Powder*, International Lead Zinc Research Organization, New York, 1982.
3. C. H. Mathewson, ed., *Zinc*, Reinhold Publishing Co., New York, 1959.
4. D. O. Rausch and B. C. Mariacher, eds., *AIIME-World Symposium on Mining and Metallurgy of Lead and Zinc*, Vol. I, American Institute of Mining Engineers, New York, 1970.
5. C. H. Cotterill and C. M. Cigan, eds., *AIIME Symposium on Mining and Metallurgy of Lead and Zinc*, Vol. II, American Institute of Mining Engineers, New York, 1970.
6. J. M. Cigan, T. S. Mackey, and T. J. O'Keefe, eds., *Lead-Zinc-Tin '80*, American Institute of Mining Engineers, Warrendale, Pa., 1979.
7. J. N. Anderson and P. E. Queneau, eds., *Pyrometallurgical Processes in Nonferrous Metallurgy*, Gordon and Breach Science Publishing, New York, 1967.
8. E. S. Dana in W. E. Ford, ed., *A Textbook of Mineralogy*, 4th ed., John Wiley & Sons, Inc., New York, 1945, p. 851.
9. A. M. Bateman in ref. 3, Chapt. 3.
10. H. Wedow and co-workers, *U.S. Geol. Sur. Prof. Pap.* **820**, 697 (1973).
11. *Minerals Yearbook*, U.S. Bureau of Mines, Dept. of the Interior, 1992.
12. J. W. Chandler in ref. 4, Chapt. 2.
13. A. L. Hatch, K. C. Hendrick, and W. Seis in ref. 5, Chapt. 2.
14. *Canadian Mines Handbook*, Northern Miner Press, Toronto, 1970–1971, p. 350.
15. E. W. Horvick in H. Baker, ed., *Metals Handbook*, 9th ed., Vol. 2, American Society for Metals, Metals Park, Ohio, 1979, pp. 824–826.
16. E. N. Doyle in ref. 4, Chapt. 40.
17. J. S. Pellett, *Mining Eng. (N.Y.)* **5**, 1211 (1953).
18. T. G. White and K. L. Clifford in D. O. Rausch and co-workers, eds., *Lead-Zinc Update*, American Institute of Mining Engineers, New York, 1977, pp. 119–147.
19. M. Rey and co-workers, *Trans. AIIME* **199**, 416 (1954).
20. T. R. A. Davey in ref. 6, pp. 48–65.
21. W. Gowland, *The Metallurgy of Non-Ferrous Metals*, 2nd ed., C. Griffen, London, 1918.
22. K. D. McBean in ref. 3, Chapt. 6.
23. **U.S. Pats. 1,884,348** (Oct. 25, 1932); **1,963,288** (June 19, 1934); **2,089,306** (Aug. 10, 1937), B. A. Stimmel, K. D. McBean, and G. Cruickshank (to Consolidated Mining and Smelting Co. of Canada).
24. J. H. Reid in ref. 7, Chapt. 5.
25. M. M. Avedesian, preprint, *Conference of the Canadian Society of Chemical Engineers, Ottawa*, Paper No. 5d, Oct. 21, 1974.
26. K. Natesan and W. O. Philbrook, *Metall. Trans.* **1**, 1353 (1970).
27. G. Stankovic in ref. 5, Chapt. 4.
28. M. Okozaki, Y. Nakane, and H. Noguci in ref. 7, Chapt. 2.
29. A. Berg and F. Pape, *Erzmetall* **31**, 224 (1978).
30. L. A. Painter in ref. 6, Chapt. 8.
31. G. Steintveit in ref. 6, pp. 85–96.
32. K. H. Heino and co-workers in ref. 5, Chapt. 5.
33. R. Denoiseux and co-workers in ref. 6, Chapt. 5.
34. R. E. Lund and co-workers in ref. 5, Chapt. 20.
35. *Eng. Min. J.* **182**(9), 99 (1981).
36. R. R. Knobler and co-workers, *Erzmetall* **32**, 109 (1979).
37. S. Tsunoda and co-workers, *AIIME-TMS Paper A73-65*, 1973.
38. F. S. Gaunce and co-workers, *CIM Bull.* **67**, 116 (1974).
39. K. Shimokawa and T. Takesue, *MMIJ-AIIME Joint Meeting Paper TIVCZ*, May 1972.
40. V. Arregui, A. R. Gordon, and G. Steintveit in ref. 6, pp. 97–123.
41. J. E. Dutrizac in ref. 6, pp. 32–64.
42. **Can. Pat. 238,439** (Oct. 22, 1975); **U.S. Pat. 4,004,991** (Jan. 25, 1977), H. Veltman and co-workers (to Sherritt Gordon Mines, Ltd.).
43. H. Veltman and G. L. Bolton, *Erzmetall* **33**, 76 (1980).
44. E. G. Parker, *CIM Bull.* **74**, 145 (1981).
45. F. S. Weimer, G. T. Weber, and R. J. Lapee in ref. 3, Chapt. 6.
46. H. H. Fukubayashi, T. J. O'Keefe, and W. C. Clinton, *U.S. Bur. Mines Rep. Invest.*, 7966 (1974).
47. G. M. Meisel, *J. Metals* **26**(8), 25 (1974).
48. D. E. Rodier in ref. 6, Chapt. 9.
49. S. Fugleberg and co-workers in ref. 6, Chapt. 10.
50. N. G. Freeman and A. Pyatt in ref. 6, pp. 222–246.
51. C. L. Mantell, *Electrochemical Engineering*, McGraw-Hill, Inc., New York, 1960.
52. R. C. Kerby and C. J. Krauss in ref. 6, pp. 187–203.
53. C. G. Maier, *U.S. Bur. Mines Bull.*, 324 (1930).
54. E. C. Truesdale and R. K. Waring, *Trans. AIIME* **159**, 97 (1944).
55. R. M. Sellwood in ref. 5, Chapt. 21.
56. H. H. Kellogg in ref. 6, pp. 28–47.
57. W. Hopkins and A. W. Richards, *J. Met.* **30**, 12 (1978).
58. C. F. Harris, A. W. Richards, and A. W. Robson in ref. 6, pp. 247–260.
59. H. Nakagawa and co-workers in ref. 6, Chapt. 16.
60. L. D. Fetterolf and co-workers in ref. 5, Chapt. 19.
61. E. L. Miller in ref. 3, Chapt. 7.
62. G. F. Weaton in ref. 3, pp. 272–273.

63. E. D. Nogueira, J. M. Regife, and A. M. Arcocha, *Eng. Min. J.* **180**(10), 92 (1979).
64. *Minerals Year Book*, U.S. Bureau of Mines, Washington, D.C., 1970–1980.
65. *Non-Ferrous Metal Data*, American Bureau of Metal Statistics, Inc., New York, 1926–1980.
66. *Lead and Zinc Statistics*, monthly bulletin of the International Lead Zinc Study Group, April 1997.
67. Ref. 65, 1970, p. 90 and 1980, p. 86.
68. Ref. 65, 1970, p. 91 and 1980, p. 87.
69. H. H. Landsberg and co-workers, *Resources in America's Future—Patterns of Requirements and Availabilities, 1960–2000*, The Johns Hopkins Press, Baltimore, Md., 1963.
70. *1981 U.S. Industrial Outlook for 200 Industries with Projections for 1985*, Bureau of Industrial Economics, U.S. Government Printing Office, Washington, D.C., Jan. 1981.
71. W. Malenbaum, *World Demand for Raw Materials in 1985 and 2000*, McGraw-Hill, Inc., New York, 1978.
72. *Annual Review, U.S. Zinc and Cadmium Industries*, Zinc Institute, Inc., New York, 1970–1980.
73. *U.S. Automotive Market for Zinc Die Casting*, Zinc Institute, Inc., New York, 1974–1980.
74. *ASTM Standards, Specification B 6-77*, American Society for Testing and Materials, Philadelphia, Pa., 1981, Part 8, pp. 11–14.
75. C. H. Mathewson in ref. 3, Chapt. 9, pp. 386–390.
76. T. G. White, *Min. Congr. J.* **63**(5), 54 (1977).
77. J. C. Van Loon, *Analytical Atomic Absorption Spectroscopy*, Academic Press, New York, 1980.
78. N. H. Furman, ed., *Standard Methods of Chemical Analysis*, Vol. I, D. Van Nostrand Co., Princeton, N.J., 1962, Chapt. 52.
79. *Annual Book of Standards*, American Society for Testing and Materials, Philadelphia, Pa., 1980, Parts 12 and 42.
80. *Zinc*, National Research Council, Subcommittee on Zinc, University Park Press, Baltimore, Md., 1979, pp. 273–296, 327–341.
81. C. T. Walsh, H. H. Sandstead, A. S. Prasad, P. M. Newberne, and P. J. Fraker, "Zinc: Health Effects and Research Priorities for the 1990s," *Environ. Health Perspect.* **102**, 5–46 (June 1994).
82. Agency for Toxic Substances and Disease Registry, *Toxicological Profile for Zinc*, May 1994.
83. IPCS Task Group on Zinc, *International Programme on Chemical Safety, Environmental Health Criteria for Zinc*, September 20, 1996.
84. *National Research Council, Subcommittee on Zinc*, University Park Press, Baltimore, Md., 1979, pp. 273–296, 327–341.
85. J. D. Nriagu, ed., *Zinc in the Environment*, John Wiley & Sons, Inc., New York, 1980, Parts 1 and 2.
86. *Zinc Coatings for Corrosion Protection*, Zinc Institute, Inc., New York, 1977, 52 pp.
87. *ASTM Standard, Specification B 240-79*, American Society for Testing and Materials, Philadelphia, Pa., 1980, Part 7, pp. 332–334.
88. *ASTM Standard, Specification B 86-76*, American Society for Testing and Materials, Philadelphia, Pa., 1980, Part 7, pp. 45–49.
89. E. W. Horvick in ref. 15, pp. 629–637.
90. *Engineering Properties of Zinc Alloys*, International Lead Zinc Research Organization, New York, 1980, pp. 1–36.
91. F. R. Sauerwine, T. P. Groeneveld, and F. C. Bennett, *Tech Data '80*, The American Die Casting Institute—Die Casting Research Institute, Des Plaines, Ill., Technical Seminar, March 20, 1980.
92. M. Hansen, *Constitution of Binary Alloys*, McGraw-Hill Book Co., Inc., New York, 1958, p. 149.
93. P. Krick in ref. 6, pp. 693–699.
94. E. A. Anderson and G. L. Werley, *Zamak Alloys for Zinc Die Casting*, 12th ed., The New Jersey Zinc Company, Palmerton, Pa., 1965.
95. G. L. Werley in T. Lyman, ed., *Metals Handbook*, 1948 ed., American Society for Metals, Cleveland, Ohio, 1948, pp. 1078–1082.
96. C. W. Roberts, *Metallurgia* **64**, 57 (Aug. 1961).
97. T. Lyman, ed., *Metals Handbook*, 8th ed., Vol. 5, American Society for Metals, Metals Park, Ohio, 1970, pp. 444–448.
98. Ref. 97, pp. 285–333.
99. A. Street, *The Die Casting Book*, Portcollis Press, Ltd., Surrey, UK, 1977.
100. T. P. Groeneveld and W. D. Kaiser, *Designing for Thin-Wall Zinc Die Casting*, International Lead Zinc Research Organization, New York, 1975.
101. W. D. Kaiser and T. P. Groeneveld, *Trans. Soc. Die Cast. Eng.* (March 1975).
102. A. J. Davis, H. Siauw, and G. N. Payne, *Trans. Soc. Die Cast. Eng.* (June 1977).
103. D. C. H. Nevison, *Using Computer Aids in Die Casting*, 10th International Pressure Die Casting Conference, May 10–14, 1981, Zinc Institute Inc., New York, 1981, 16 pp.
104. D. C. H. Nevison, *Precis. Met.* **32**(3), 27 (March 1974).
105. D. C. H. Nevison, *Ind. Design* **21**(7), 51 (Sept. 1974).
106. E. T. Foster, *Appliance Engineer* **9**, 67 (April 1975).
107. F. R. Sauerwine and A. D. Behler, *10th Society of Die Casting Engineers Congress*, Paper No. G-T79-101, St. Louis, Mo., March 1979, 16 pp.
108. J. T. Winship, *Am. Mach.* (Aug. 1979).
109. *Design Applications of Gravity Cast Zinc*, International Lead Zinc Research Organization Zinc Institute, New York, 1972, 36 pp.
110. E. Gervais, H. Levert, and M. Bess, *84th Casting Congress and Exposition of the American Foundrymen's Society*, St. Louis, Mo., April 21–25, 1980, American Foundrymen's Society, Des Plaines, Ill.
111. *ASTM Standards, Specification B 669-80*, American Society for Testing and Materials, Philadelphia, Pa., 1981, Part 8, pp. 789–792.
112. D. Apelian, M. Paliwal, and D. C. Herrschaft, *J. Met.*, 12 (Nov. 1981).
113. E. Gervais, A. Y. Kandeil, and H. Levert, *Die Cast Eng.* **25**, 44 (Sept.–Oct. 1981).
114. *Zinc Foundry Alloys*, Zinc Institute, Inc., New York, 1980, 8 pp.
115. W. Mihaichuk, *Mod. Cast.* **71**, 39 (July 1981).
116. T. Riedlinger, *Mod. Met.* **37**, 12, Chicago, Ill. (Jan. 1981).
117. J. W. Kissel and K. F. Dufrane, *The Bearing Properties of Zinc Gravity Casting Alloys*, ILZRO Project ZM-298, Progress Report No. 2, International Lead Zinc Research Organization, New York, 1980.
118. G. L. Werley in ref. 95, p. 334.
119. J. C. Fox in ref. 3, Chapt. 9, pp. 423–453.
120. Ref. 97, pp. 222–236, 265–284.
121. T. Lyman, ed., *Metals Handbook*, 8th ed., Vol. 4, American Society for Metals, Metals Park, Ohio, 1969, pp. 145–153.
122. *ASTM Standard, Specification B 69-66*, reapproved 1979, American Society for Testing and Materials, Philadelphia, Pa., 1981, Part 8, pp. 35–40.
123. Ref. 90, Chapt. 3, pp. 57–60.
124. R. K. Martin in ref. 3, Chapt. 9, pp. 523–527.
125. E. H. Kelton in ref. 95, pp. 1082–1084.
126. C. W. Roberts and B. Walters, *J. Inst. Met.* **76**, 557 (1949–1950).
127. E. W. Horvick in ref. 15, pp. 629–638.
128. S. W. K. Morgan, *Zinc and Its Alloys*, MacDonald and Evans, Ltd., London, UK, 1977, p. 155.
129. *Steel (USA)* **161**(3), 83 (July 17, 1967).
130. B. D. Wakefield, *Iron Age* **204**(21), 122 (Nov. 20, 1969).
131. **U.S. Pat. 3,865,176** (Feb. 11, 1975), J. Dampas and R. W. Hazelett (to Hazelett Strip Casting Corp.).

132. W. J. Mosbach, *Met. Prog.*, 60 (May 1975).

133. *Plating and Surface Finishing* **68**(7), 36 (July 1981).

134. A. Collart and C. P. Tombras, Sr., *Ind. Research and Development* **24**(2), 170 (Feb. 1982).

135. *Superplastic Metal Manual*, The New Jersey Zinc Company, Palmerton, Pa., 1970.

136. R. H. Johnson, *Met. Mater.* **4**(9), 115 (1970).

137. C. B. Tennant, *Paper No. A70-48, TMS-AIME Annual Meeting*, Denver, Colo., Feb. 1970.

138. W. Showak and S. R. Dunbar in T. Lyman, ed., *Metals Handbook*, 8th ed., Vol. 7, American Society for Metals, Metals Park, Ohio, 1972, pp. 335–340.

139. D. R. Dreger, *Mach. Des.*, 110 (Aug. 8, 1974).

140. J. N. Pomeroy and co-worker in ref. 3, Chapt. 6, pp. 314–321.

141. A. Pass, *J. Oil and Colour Chem. Assoc.* **35**, 241 (June 1952).

142. D. S. Newton in ref. 5, pp. 99–1005.

143. *ASTM Standards, Specification D 520-51*, reapproved 1976, American Society for Testing and Materials, Philadelphia, Pa., 1981, Part 28.

144. U. R. Evans, *Metal Ind. (London)* **67**, 114, 118 (1945).

145. J. E. O. Mayne and U. R. Evans, *J. Soc. Chem. Ind. London Rev. Sec.* **22**, 109 (1944).

146. A. W. Kennedy, *Modern Paint and Coatings* **66**(9), 21 (1976).

147. A. C. Elm, *Zinc Dust Metal Protective Coatings*, The New Jersey Zinc Company, Palmerton, Pa., 1968.

148. C. G. Munger, *Mater. Prot.* **2**(3), 8 (March 1963).

149. C. T. Flachbarth and W. H. Taylor in ref. 3, Chapt. 9, pp. 511–513.

Frank E. Goodwin
International Lead and Zinc Research Organization, Inc.



ZINC COMPOUNDS

Zinc usually occurs as the sulfide but significant quantities of the oxide, carbonate, silicate [14374-77-7], and basic compounds of the latter two are also mined (see ZINC AND ZINC ALLOYS).

Properties

Zinc is in Group IIB of the periodic table and exhibits a valence of +2 in all its compounds. Being high on the electromotive series, zinc forms quite stable compounds and, as such, resembles magnesium. Bonding in zinc compounds tends to be covalent, as in the sulfide and oxide. With strongly electropositive elements, eg, chlorine, the bond is more ionic. Zinc also tends to form stable covalent complex ions, eg, with ammonia $[Zn(NH_3)_2]^{2+}$, cyanide $[Zn(CN)_4]^{2-}$, and hydroxyl $[Zn(OH)_4]^{2-}$. The coordination number is usually four, to a lesser degree six, and in some cases five. A good review of zinc compounds is given in Ref. 1.

Zinc forms salts with acids but since it is amphoteric, it also forms zincates, eg, $[Zn(OH)_3 \cdot H_2O]$ and $[Zn(OH)_4]^{2-}$. The tendency of zinc to form stable hydroxy complexes is also important because some basic zinc salts are only slightly soluble in water. Examples are $3Zn(OH)_2 \cdot ZnSO_4$ [12027-98-4] and $4Zn(OH)_2 \cdot ZnCl_2$ [11073-22-6], which may precipitate upon neutralization of acidic solutions of the salts.

Properties of zinc salts of inorganic and organic salts are listed in Table 1 with other commercially important zinc chemicals. In the dithiocarbamates, 2-mercaptobenzothiazole, and formaldehyde sulfoxylate, zinc is covalently bound to sulfur. In compounds such as the oxide, borate, and silicate, the covalent bonds with oxygen are very stable. Zinc-carbon bonds occur in diorganozinc compounds, eg, diethylzinc [557-20-0]. Such compounds were much used in organic synthesis prior to the development of the more convenient Grignard route (see GRIGNARD REACTIONS).

The water solubility of zinc compounds varies greatly, as shown in Table 1. Water-soluble compounds not listed are zinc formate [557-41-5] chlorate [10361-95-2], fluorosilicate [16871-71-9], and thiocyanate [557-42-6]. Also, the water-soluble amino and cyanide complexes have many uses.

Zinc compounds are generally colorless unless the other component, eg, chromate, is colored. The lack of color of most zinc compounds in visible light is a great advantage in that they do not color paint films, plastics, rubber, cosmetics, etc. However, when excited by various types of radiation and at various temperatures, zinc oxide, sulfide, selenide [1315-09-9], and related compounds exhibit luminescence, ie, they emit colored light (see LUMINESCENT MATERIALS). Zinc-based phosphors can be produced in many colors, depending upon the added dopants. They are used in television tubes, luminescent glasses, and various specialty products.

Zinc Oxide

Physical Properties. Some of the physical properties of zinc oxide are listed in Table 2. Of great importance is the fact that it completely absorbs ultraviolet light below 366 nm and, thus, is unique among white pigments (15). Its high refractive indexes make it a good white pigment where its mean diameter for maximum light scattering is 0.25 μ m. The crystal structure of zinc oxide is likely to stabilize defects, eg, zinc excess or deficiency and inclusion of foreign ions, and therefore has useful semiconductor properties. Variously doped oxides are used for photocopying, catalysts, and phosphors (16,17) (see CATALYSTS; SEMICONDUCTORS).

Table 1. Properties, Prices, and Uses of Zinc Compounds

Zinc compound	Formula, synonym	CAS Registry No.	Sp gr	Mp, °C	Solubility ^a , g/100 g solvent			Price ^b , \$/kg (Aug. 1981)	Uses
					Water	Other			
acetate	Zn(C ₂ H ₃ O ₂) ₂ ·2H ₂ O	[5970-45-6]	1.735	237	40 ^{25°} C	67 ^{100°} C	3 ^{25°} C alcohol	3.53	wood preservative, mordant antiseptics, catalyst, waterproofing
ammonium chloride	ZnCl ₂ ·2NH ₄ Cl	[52628-25-8]	1.88	150 dec	66 ^{0°} C	69 ^{10°} C		0.71	galvanizing, solder flux, adhesives
diborate	ZnOB ₂ O ₃ ·2H ₂ O	[27043-84-1]	3.64		0.007 ^{25°} C		sl sol HCl	1.21	fireproofing, ceramics, fungicide
dodecaborate	2ZnO·3B ₂ O ₃ ·3.5H ₂ O	[12513-27-8]	4.22	980	insol			1.54	fire retardant
bromide	ZnBr ₂	[13550-22-6]	4.21	394	471 ^{25°} C	675 ^{100°} C	sol alcohol, ether	62.00 ^c	photographic paper, catalyst, batteries
carbonate	ZnCO ₃	[3486-35-9]	4.40	CO ₂ at 300	0.001 ^{15°} C		insol alcohol	45.40 ^c	ceramics, rubber, astringent (lotions)
chloride	ZnCl ₂	[7646-85-7]	2.91	275	432 ^{25°} C	614 ^{100°} C	sol ether	0.41	textiles, adhesives, flux, wood preservative, antiseptic, astringent
cyanide	Zn(CN) ₂	[557-21-1]	1.85	800 dec	0.005 ^{20°} C		sol alkali, CN	3.63	electroplating, gold extraction
dithiocarbamates dibutyl	$Zn \left[\begin{array}{c} S \\ \\ SCN(C_4H_9)_2 \end{array} \right]_2$	[136-23-2]	1.21	106	insol		sol C ₂ H ₅ , CS ₂ , CHCl ₃	3.17	vulcanization accelerator, lube oil
diethyl	$Zn \left[\begin{array}{c} S \\ \\ SCN(C_2H_5)_2 \end{array} \right]_2$	[14324-55-1]	1.48	176	insol		sol C ₂ H ₅ , CS ₂ , CHCl ₃	3.32	vulcanization accelerator

ethylenebis	$\text{Zn} \left[\begin{array}{c} \text{S} \\ \parallel \\ \text{SCNHCH}_2 \end{array} \right]_2,$	[12122-67-7] /			insol		sol C H , CS ₂ , CHCl ₃	5.50	fungicide, insecticide
dimethyl	zineb $\text{Zn} \left[\begin{array}{c} \text{S} \\ \parallel \\ \text{SCN}(\text{CH}_3)_2 \end{array} \right]_2,$	[137-30-4] /	1.71	249	0.0065 ^{25° C}		sol CS ₂ , ace-tone, alkali	3.06	vulcanization accel-erator, fungicide
2-ethylhexanoate	ziram Zn(C ₈ H ₁₇ O ₂) ₂	[136-53-8] /	0.90		insol		sol hydro-carbo n	1.87	paint drier, silicone rubber cure
fluoroborate	Zn(BF ₄) ₂ ·6H ₂ O	[13826-88-5] /		H ₂ O at 60°C	>100 ^{25° C}		sol alcohol	0.60	plating, bonderizing, textile resin cure
fluoride	ZnF ₂	[7783-49-5] /	4.95	872	1.61 ^{18° C}		sol hot acid, NH ₄ OH	100.0	ceramics, impregnating wood, galvanizing
formaldehyde sulfoxylate	Zn(HSO ₂ -CH ₂ O) ₂	[24887-06-7] /		90 dec	60 ^{25° C}		insol alcohol	2.31	reducing agent, drying, polymerization
hydrosulfite	ZnS ₂ O ₄ ·2H ₂ O	[7779-86-4] /		200 dec	40 ^{20° C}			2.25	bleach, especially textile, paper, reducing agent
iodide	ZnI ₂	[10139-47-6] /	4.70	446	432 ^{18° C}	510 ^{100° C}	sol alcohol	182 ^c	medicine, photography
2-mercaptobenzo-t hiazole	Zn(SC H ₄ NCS) ₂	[155-04-4] /	1.70	300 dec	insol	insol	sol C H , CS ₂ , CHCl ₃	3.17	vulcanization accelera-tor for latex
naphthenate	Zn[(C ₂ H ₅) ₅ CHCOO I ₂	[12001-85-3] /			insol	insol	sol hydrocar-bo n, acid	1.17	paint film improver, rot proofer
nitrate	Zn(NO ₃) ₂ ·6H ₂ O	[10196-18-6] /	2.06 5	18	93	900 ^{70° C}	sol alcohol	0.75	textiles as resin cata-lyst, mordant
oxide	ZnO	[1314-13-2] /	5.47, 5.61	1800 sub-lim es	0.00042 ^{18° C}		sol acid, alkali, NH ₄ OH	1.05– 1.12	latex coagulant vulcanization accelera-tor, mildewstat, pig-ment, supplement in feed and fertilizer, catalyst, ceramics, intermediate
peroxide	ZnO ₂	[1314-22-3] /	1.57	212 ex-plod es	insol		sol acid	5.50	cosmetic powders as antiseptic
phosphate	Zn ₃ (PO ₄) ₂	[7779-90-0] /	4.00	900	2.6 ^{25° C}	insol	sol NH ₄ OH, insol alcohol	1.85	metal coatings, dental cement
potassium chromate	4ZnO·K ₂ O·4CrO ₃ ·3H ₂ O, zinc yellow	[37300-23-5] /	3.36 –3.4 6		0.24 ^{25° C}			2.47	rust-inhibiting pigment
resinate	Zn(C ₂₀ H ₃₀ O ₂) ₂	[9010-69-9] /	1.24	205	insol	insol	sol hydro-carbo n	0.99	inks, paint drier
selenide	ZnSe	[1315-09-9] /	5.42	>1100	insol		sol acid	330.0	phosphor
silicofluoride	ZnSiF ₆ ·6H ₂ O	[18433-42-6] /	2.10	100 dec	77 ^{10° C}	93 ^{50° C}		0.37	laundry sour, wood preservative, plaster additive
stearate	Zn(C ₁₇ H ₃₅ COO) ₂	[557-05-1] /	1.09	120	insol	insol	sol hydro-carbo n	1.90	lubricant, mold release, vinyl stabilizer, anti-cake, water repellant
sulfate	ZnSO ₄	[7733-02-0] /	3.54	680 dec	41.9 ^{0° C}	91 ^{70° C}	sol glycerol	0.77	rayon bath, agriculture, zinc plating, interme-diate,
sulfate	ZnSO ₄ ·H ₂ O	[7446-19-7] /	3.28	238 dec	101 ^{70° C}	87 ^{105° C}		0.58	flotation, mordant rayon bath, agriculture, zinc plating, interme-diate,
sulfide	ZnS	[1314-98-3] /	3.98, 4.10	1185 sub-lim es	0.0007 ^{18° C}	insol	sol acid	1.76 ^d	flotation, mordant phosphor, white pig-ment, dental materials
tetroxychromate	4Zn(OH) ₂ ·ZnCrO ₄	[13530-65-9] /	3.87 –3.9 7		0.01 ^{25° C}			2.57	wash primer
undecylenate	Zn[CH ₂ – CH(CH ₂) ₈ ·COO] ₂	[557-08-4] /		115	insol	insol	sol hydro-carbo n	9.70	dermal fungicide

^a Insol insoluble, v sol very soluble, dec decompose.
^b Ref. 2.
^c Reagent price.
^d Pigment price.

Table 2. Selected Physical Properties of Zinc Oxide

Property	Value	Ref.
mp, °C	ca 1975 (subl)	3
color	white in finely divided form	
refractive index, 0.5 μm	2.015, 2.068	4
specific gravity	5.68	
water solubility, minimum at pH	9.7	5
K_{sp} Zn(OH) ₂	4.5×10^{-17}	
heat capacity (at 25°C), J (mol°C) ^a	40.26	3
$\Delta H_{\text{formation}}$ (at 419.5–907°C), kJ mol ^a	356.1	6
$\Delta F_{\text{formation}}$ (at 419.5°C), kJ mol ^a	281.6	6
$\Delta F_{\text{formation}}$ (at 907°C), kJ mol ^a	–229.0	
$\mathcal{S}$ formation (at 25°C), J mol ^a	43.65	7
coefficient of expansion, $\times 10^{-6}$ °C	4.0	8
conductivity, W (m·K)	25.2	9
crystal structure	hexagonal, wurtzite	
conductivity (<i>n</i> -type), S cm	$10^{-7} - 10^3$	10
piezoelectricity (lithium-doped)	ca 4× that of quartz	11
magnetic susceptibility (at 196°C), $\times 10^{-6}$ Hz units	0.20	12
pyroelectric current density, MA (m ² ·sK)	6.8	13
E° of Zn + ¹ / ₂ O ₂ → ZnO (at 25°C), V	1.649	14

^a To convert J to cal, divide by 4.184.

Chemical Properties. Zinc oxide, as an amphoteric material, reacts with acids to form zinc salts and with strong alkalies to form zincates. In the vulcanization of rubber, the chemical role of zinc oxide is complex and the free oxide is required, probably as an activator (see RUBBER,CHEMICALS). Zinc soaps result from the oxide's interaction with organic acids and their interaction with the accelerator. The oxide is used alone as an accelerator in certain elastomers, eg, neoprene and thiokols, which contain chlorine and sulfur in the polymer molecules. Zinc oxide reacts with carbon dioxide in moist air to form oxycarbonate. Acidic gases, eg, hydrogen sulfide, sulfur dioxide, and chlorine, react with zinc oxide, and carbon monoxide or hydrogen reduce it to the metal. At high temperatures, zinc oxide replaces sodium oxide in silicate glasses. An important biochemical property of the oxide is its fungicidal moldewstatic action (see FUNGICIDES) (18). It is also soluble in body fluids and soils (19).

Production and Processing. Primary zinc oxide is manufactured by oxidizing zinc vapor in burners wherein the concentration of zinc vapor and the flow of air is controlled so as to develop the desired particle size and shape. The hot gases and particulate oxide or fume pass through tubular coolers and then the ZnO is separated in a baghouse. The purity of the zinc oxide depends upon the source of the zinc vapor. Zinc oxide of great purity is required for pharmaceutical, photoconductive, and certain other grades, and these are made by the indirect (French) process which accounted for 34% of U.S. zinc production in 1976–1980. Zinc vapor from previously purified zinc metal is burned (20,21). Less-pure zinc oxide is manufactured by the direct (American) process, by which impure zinc oxide is reduced to zinc vapor which is then burned (20). The American process accounted for 58% of U.S. zinc oxide production in 1976–1980. Certain impurities in the original crude ore and carbonaceous reductant are inseparable from the product oxide. Nevertheless, the uses dictate that certain impurities be at low levels. For instance, cadmium, lead, iron, sulfur, copper, and manganese can be deleterious in rubber if they exceed certain concentrations (22,23). Some scrap zinc materials undergo both French and American processes; other processes are used to produce secondary zinc oxide from zinc scrap, zinc-containing sludges, and metallurgical slags.

Direct (American) Process.

Grate Furnaces. In the eastern Wetherhill furnace, four or more firebrick furnaces (called a block) having common walls are charged in cyclic fashion. Coal that is hot from the previous charge is first spread on the grate and, after ignition, a damp, well-blended mixture of zinciferous material and coal is added. The bed is maintained in a reducing condition with carbon monoxide to produce zinc and lead, if present. Metal vapors are drawn into a chamber above the furnace, where combustion air oxidizes them to pigment. The hot pigment–gas stream enters a cooling duct common to the whole block and, in this way, the product becomes a uniform blend. The raw material contains 15–75 wt % zinc, but usually over 60 wt %, and the coal is adjusted accordingly. For best results, anthracite or coke is used (see COAL). Calcium, iron, and silica influence the nature of the clinker and the elimination of zinc and must, therefore, be minimized.

The western Wetherill furnace is like the eastern furnace except that a block of 12 or more furnaces has a common combustion chamber. This results in improved control and uniformity of product. Clinker usually contains 8–15 wt % zinc, which is often recovered in a Waelz kiln (24).

The traveling-grate furnace requires less labor, increases the output per unit of grate area, and produces more uniform product than the Wetherill furnaces. The traveling grate is an endless chain of cast-iron bars, driven by sprockets, which traverses a firebrick chamber. Anthracite briquettes are fed to a depth of ca 15 cm. After ignition by the previous charge, the coal briquettes are covered by 15–16.5 cm of ore coal briquettes. The latter are dried with waste heat from the furnace. Zinc vapor evolves and burns in a combustion chamber and the spent clinker falls into containers for removal (24,25).

Rotary Kilns. The pigment-grade oxide kiln, because of its high temperature, produces pigment-quality zinc oxide and makes possible higher recovery as compared to the grate furnaces. The kilns, which are 2.4-m dia and 12.1–15.2-m long, are fed in the firing end with a mixture of 65 wt % sinter and 35 wt % anthracite. Very slow rotation and small slope give sufficient retention time for good zinc elimination. The atmosphere in the kiln is reducing. Air is then control-fed into a firebrick combustion chamber, where the zinc is burned to the desired size and shape. The exiting solid residue is water-quenched.

The Waelz process is used to beneficiate many metals, including zinc, and produces a relatively impure oxide (26). This is true because the loose feed is charged at the end of the kiln where metal vapors are exiting and because the feed materials are usually quite impure. Where the process is used to make crude zinc oxide, feeds include oxidized and sulfide zinc ores, residues, zinc-bearing iron ores and flue dusts, lead furnace slags, mill slimes, electrolytic-zinc leach residues, and ores of various other metals, eg, tin and gold (27). The metallurgy of the process is similar to the preceding processes, but it is complicated by the pressence of other materials in the feed, eg, sulfates, sulfides, carbonates, and silicates. Conditioners, eg, silica, lime, and iron oxide, are often added to control the fluidity of the charge (28). Also, lime can be used to fix sulfur as calcium sulfide, so that materials containing up to 20 wt % sulfur can be introduced into the kiln. Any sulfur dioxide formed tends, because of the high temperature and long retention time, to be oxidized to trioxide and reacts with metallic oxides to form sulfate.

The charge, which contains ca 25 wt % carbon, progresses through the 11–30-m long kiln, countercurrent to the hot gases from the fuel burner and the burning carbon in the bed. Lifters give the bed a considerable mixing action, as the kiln rotates at 1–1.5 rpm. The particles are reduced when submerged in the bed and are oxidized on the surface. Conditions are adjusted so that the products in the exiting gas are zinc and cadmium vapors, stannous oxide, germanium monoxide, carbon monoxide, and often further oxidized states of these compounds. They are fully oxidized in the combustion chamber which may also serve as a settling chamber for fly ash from the kiln. Settled material is recycled to the kiln. The fume is cooled and collected in a baghouse. Very often, the crude ZnO is sintered to remove cadmium, sulfur, and lead and used for the production of zinc oxide or zinc metal.

Electrothermic Process. Electrothermic zinc smelting is described in refs. 29–30. The oxide furnaces in the one U.S. plant are 11-m high, three are 1.75-m dia,

and the fourth is 2.44-m dia with eight graphite electrodes at each of two levels. The electrical resistance path through the descending charge of 70 wt % sinter and 30 wt % coke is 9-m long, and it supplies the energy for smelting. The zinc vapor and carbon monoxide pass through exit ports at four levels between the upper and lower electrodes into manifolds where pigment forms by oxidation with added air. After cooling and passing through a cleaning cyclone, the oxide is collected in a baghouse (31).

Indirect (French) Process. Zinc metal vapor for burning is produced in several ways, one of which involves horizontal retorts. Since all the vapor is burned in a combustion chamber, the purity of the oxide depends on that of the zinc feed. Oxide of the highest purity requires special high grade zinc and less-pure products are made by blending in Prime Western and even scrap zinc.

Another method involves an electric-arc vaporizer which is >2000°C before burning (25,32). One of the features of the process is a rapid quench of the hot gas flow to yield very fine oxide particles (<0.15 nm). This product is quite reactive and imparts accelerated cure rates to rubber. Internally fired rotary kilns are used extensively in Canada and Europe and, to a limited extent, in the United States (24). The burning occurs in the kiln and the heat is sufficient to melt and vaporize the zinc. Because of the lower temperatures, the particles are coarser than those produced in the other processes. In a fourth process, zinc metal which is purified in a vertical refining column is burned. In essence, the purification is a distillation and impure zinc can be used to make extremely pure oxide. Also, a wide range of particle sizes is possible (33).

Secondary Zinc Oxide. Secondary oxide is that made from scrap, eg, galvanizer's dross, trimmings, automotive zinc scrap, etc, and is largely used by rubber and ceramics manufacturers. About one-fifth of the oxide produced in the United States is from secondary material, although some of it is made by primary producers. Most secondary oxide is of the French type in that the starting zinc is metallic. The zinc is usually vaporized in upright retorts in such a way that the higher boiling impurities, eg, lead, do not vaporize and relatively pure oxide is produced (34). Another procedure for manufacturing secondary oxide is precipitation of zinc carbonate from waste zinc solution, followed by drying, calcining, and grinding to a coarse oxide suited for ceramics. A similar product results from the calcination of zinc hydrosulfite [7779-86-4] sludge. Such sludges are by-products of the use of zinc hydrosulfite as a bleaching agent.

Leaded Zinc Oxide. Oxides containing more than 5 wt % basic lead sulfate are classified as leaded and are made in the American process from high lead materials, usually lead sulfide mineral, or by blending zinc oxide and basic lead sulfate. There is only one manufacturer in the United States and the product contains 20–28 wt % basic lead sulfate. Leaded oxides are used only in rubber in the United States.

Slag-Fuming Process. Three plants in the United States, two in Canada, and one in Mexico process slag from blast furnaces (usually lead) and residues to make zinc oxide fume. The oxide is impure and most is sent to zinc smelters for conversion into metal; however, a small amount is sold as pigment. The process consists of blowing powdered coal into a bath of molten slag via multiple double-inlet tuyeres (35). Zinc is eliminated from the 10–18 wt % zinc charge in ca 2 h, and the coal is the fuel and the reductant, ie, carbon monoxide, source. The direct method of introduction of the coal combines drying and grinding with the feeding of coal into the furnace feed lines. Indirect feeding uses previously dried and ground coal from a bin. In either case, the coal–air mixture must be carefully controlled to minimize excess coal yet avoid too low a temperature. The slag temperature is maintained at 1150–1200°C. Recovery is ca 90% or better for both lead and zinc, and results in fume of ca 70 wt % zinc oxide and 7–10 wt % lead.

Zinc Oxide Treatments. There are a number of treatment procedures in addition to the customary screening and pulverizing. Various chemical materials are added to improve ease of incorporation and dispersibility of the oxide in rubber and paint (see PAINT). Commonly, these are fatty acids, eg, propionic and stearic, and oils (36,37). Phosphoric acid is added to reduce chemical activity and trialkyl phosphates to combat dusting (38,39). American-process zinc oxide may contain acidity in the form of oxysulfur compounds; this can be removed by washing with solutions of ammonia or ammonium salts (40). However, improved American-process oxides, which do not need washing, are being produced. Unless recalcined, zinc oxide tends to be bulky, dusty, and have poor dry flow. These characteristics are especially objectionable to those who use the oxide in ceramics and rubber. Therefore, some grades are pelleted in rotating drums or densified by passing through rolls.

Economic Aspects. Table 3 shows that rubber production is the largest market for zinc oxide; the downturn in 1980 resulted from a drop in tire production because of the production trend to smaller tires, more importation of tires, and a recession. The drop in paint usage reflects the trend to water-base paints, which originally contained no zinc oxide. However, its growing use in such paints is based upon improved formulations based on zinc oxide. The increased use in agriculture is a result of the realization of the importance of zinc as a trace element. The rise in use of zinc-oxide-coated paper for photocopying is followed by a slackening in use because of a shift to plain-paper copiers.

Table 3. U.S. Zinc Oxide Shipments by Industry, Metric Tons^a

Industry	1970	1975	1978	1979	1980	1995
rubber	101,079	87,279	97,989	93,075	61,796	68,200
paints	19,862	9,994	13,237	12,503	12,165	3,840
ceramics	8,175	5,715	9,245	9,236	5,702	2,710
chemicals	17,631	15,916	27,057	27,710	17,551	24,000
agriculture	2,038	1,676	4,847	4,397	6,930	w ^b
photocopying	28,894	22,359	19,096	16,148	9,604	w ^b
other	15,809	10,815	9,981	16,700	22,028	5,330
Total	193,488	153,754	181,452	179,769	135,776	104,000

^a Ref. 41.
^b w withheld to avoid proprietary disclosure included in "other".

Zinc oxide prices are listed in Table 4.

Table 4. Year-End U.S. Prices of Zinc Oxide, \$/kg^a

Grade	1975	1978	1979	1980	1996
American process, lead-free	0.89	0.90	0.98	0.97	
French process, lead-free					
lead-free	0.91	0.93	1.01	1.01	1.87
high purity	0.95	0.97	1.04	1.05	1.96
electrophotographic	0.97	0.99	1.07	1.05	2.53
leaded ^b					
12 wt % lead	0.78	0.82	0.87	0.87	
18 wt % lead					
35 wt % lead					

^a Ref. 42.
^b As basic lead sulfate.

Health and Safety Factors. Zinc oxide is considered nontoxic, but inhalation of freshly formed fume can cause zinc chills also known as brass-founder's ague. The symptoms are fever and cough followed by chills after ca 4–8 h (43). No aftereffects have been noted and workers who are

continually exposed quickly develop a resistance (44). So-called chronic zinc poisoning is caused by toxic impurities, eg, lead, cadmium, aresnic, and antimony, which commonly contaminate zinc ores (45). The ACGIH has classified zinc oxide as a nuisance particulate and set the threshold limit value for air-borne oxide at 5 mg m⁻³ (46).

Uses. The uses of zinc oxide can be divided into two groups based on the chemical and physical properties of the compound (47). The largest user, the rubber industry, uses it chemically as a vulcanization activator and accelerator and to slow rubber aging by neutralizing sulfur and organic acids formed by oxidation. Fine oxides are used for fast cures and coarse, sulfated grades for slow cures. Physically, it is a reinforcing agent, a heat conductor, a white pigment, and an absorber of uv light.

In paints, zinc oxide serves as a mildewstat and acid buffer as well as a pigment. The oxide also is a starting material for many zinc chemicals. The oxide supplies zinc in animal feeds and is a fertilizer supplement used in zinc-deficient soils. Its chemical action in cosmetics (qv) and drugs is varied and complex but, based upon its fungicidal activity, it promotes wound healing. It is also essential in nutrition. Zinc oxide is used to prepare dental cements in combination with eugenol and phosphoric and poly(acrylic acid)s (48) (see DENTAL MATERIALS).

Zinc oxide functions in ceramics in several ways. Added to glasses, it imparts low thermal expansion, low melting, and increased chemical resistivity. The semiconducting property of a great variety of glasses and ceramics is based on their zinc oxide content. Zinc ferrites are basically zinc–ferrite oxide spinels, which are highly magnetic. Usually they also contain other oxides, eg, nickel and manganese oxides. Ferrites (qv) are used in many electrical and electronic devices. The oxide is used as a catalyst in alkylation, oxidation, hydrogenation, and dehydrogenation. The oxide is also used in coated photocopy paper (see ELECTROPHOTOGRAPHY).

Nutrition. Zinc is essential to the proper functioning of plants and animals and, as zinc sulfate and oxide, it is used as a feed supplement (49–51) (see MINERAL NUTRIENTS; FEEDS AND FEED ADDITIVES). Most crops use less than a kilogram of zinc per 1000 m² per year, so that zinc salts added at 1.3–4.5 kg ha gradually build up the zinc reserve (52). Animals, including humans, store relatively little available zinc and, thus, require a constant supply in the diet. For instance, beef cattle require 10–30 mg kg dry feed, dairy cattle 40 mg kg, and breeding hens 65 mg kg. Zinc from plants is considered less available to monogastric animals than zinc from animal protein.

Zinc is second only to iron as a trace metal in humans. A 70-kg human body has ca 4.0 g iron, 2.0 g zinc, 0.2 g manganese, 0.1 g copper, and <0.1 g of all other elements combined (53). The recommended dietary allowance of zinc is 15 mg d for adults and 25 mg d for lactating females (54). Supplemental zinc is administered in the form of sulfate, chloride, acetate, gluconate [4468-02-4], stearate, or oxide but the anion is not involved in the processes of utilization. Supplementary zinc is prescribed for zinc deficiency and certain disorders. Deficiency is thought to be more widespread in the United States than previously assumed (55). Although excessive zinc produces toxic symptoms, such symptoms rarely occur. Large doses or continued ingestion of zinc may cause gastrointestinal distress; for example, 2 g of zinc sulfate is recommended as an emetic. However, 18 patients given 660 mg zinc daily for 16–26 weeks showed no ill effects but a 16-year-old boy who ingested 12 g of elemental zinc in two days did develop noticeable ill effects (56).

Insufficient zinc results in slowed growth, delayed wound healing, poor appetite, mental lethargy, and sexual immaturity and it interferes with the immune response. The main function of zinc in metabolism is enzymatic and there is evidence of other physiologic roles, eg, in stabilization of membrane structure (57).

Zinc Chloride

Zinc chloride melts at 275°C, boils at 720°C, and is stable in the vapor phase up to 900°C. It is very hygroscopic, extremely water-soluble, and soluble in organic liquids such as alcohols, esters, ketones, ethers, amides, and nitrides. Hydrates with 1, 1.5, 2.5, 3, and 4 molecules of water have been identified and great care must be exercised to avoid hydration of the anhydrous form. Aqueous solutions of zinc chloride are acidic (pH 1.0 for 6 M) and, when partially neutralized, can form slightly soluble basic chlorides, eg, ZnCl₂·4Zn(OH)₂ [11073-22-6] and Zn(OH)Cl [14031-59-5]. Many other basic chlorides have been reported (58).

Anhydrous zinc chloride can be made from the reaction of the metal with chlorine or hydrogen chloride. It is usually made commercially by the reaction of aqueous hydrochloric acid with scrap zinc materials or roasted ore, ie, crude zinc oxide. The solution is purified in various ways depending upon the impurities present. For example, iron and manganese precipitate after partial neutralization with zinc oxide or other alkali and oxidation with chlorine or sodium hypochlorite. Heavy metals are removed with zinc powder. The solution is concentrated by boiling, and hydrochloric acid is added to prevent the formation of basic chlorides. Zinc chloride is usually sold as a 47.4 wt % (sp gr 1.53) solution, but is also produced in solid form by further evaporation until, upon cooling, an almost anhydrous salt crystallizes. The solid is sometimes sold in fused form.

The fumes of zinc chloride are highly toxic and can damage mucous membranes and cause pale gray cyanation. It can also ulcerate the skin of workers using it as a soldering flux or those handling wood impregnated with it (59).

The largest use of zinc chloride in the United States is in wood preservation, fluxes, and batteries (see BATTERIES). Zinc chloride solution dissolves vegetable fiber and is widely used in mercerizing cotton (qv), swelling fibers, as a mordant in dyeing, parchментizing paper, etc (see FIBERS, VEGETABLE; DYES, APPLICATION AND EVALUATION; PAPERMAKING ADDITIVES). It dissolves metal oxides and is used as a flux, especially in galvanizing (see SOLDERS). Zinc electroplating is often done with a chloride bath (see ELECTROPLATING). In medicine, it is used in antiseptics, deodorants, dental cements, and disinfectants (qv). Zinc chloride solutions preserve wood and textiles and are used in adhesives (qv) and embalming fluids (see TEXTILES). Other uses are in organic synthesis, eg, in the preparation of methyl chloride and diethylzinc, as a dehydrant, in rubber vulcanization, and in oil refining. The consumption of zinc chloride is declining, as shown in Table 5.

Zinc Sulfate

Anhydrous zinc sulfate forms when its hydrates are heated above 238°C. At ca 680°C, sulfur trioxide separates from the compound, forming 3ZnO·2SO₃ [12037-14-8] and above 930°C the compound is decomposed to zinc oxide. The three stable hydrates are ZnSO₄·H₂O [7446-19-7], ZnSO₄·6H₂O [13986-24-8], and ZnSO₄·7H₂O (orthorhombic) [7446-20-0]. The latter heptahydrate occurs in a few small deposits as the mineral goslarite [7446-20-0]. Three unstable hydrates are ZnSO₄·4H₂O [33309-49-8], ZnSO₄·2H₂O [80867-26-1], and ZnSO₄·7H₂O (monoclinic).

Table 5. U.S. Production and Importation of Zinc Oxide, Sulfate, and Chloride, Metric Tons^a

Zinc compound	1970		1975		1980		1992	
	Produc-tion	Import	Produc-tion	Import	Produc-tion	Import	Produc-tion	Import
zinc oxide	202,059	10,952	150,050	11,963	145,509	29,843	103,037	38,997
zinc sulfate	48,236	5,713	21,223	2,895	35,159	3,871	27,763	3,828
zinc chloride ^b	20,139	1,044	na ^c	696	11,676	1,008	6,309	3,096

^a Ref. 41.
^b Includes zinc chloride in zinc ammonium chloride and chromated zinc chloride.
^c Na not available.

The solubility of zinc sulfate increases almost linearly with temperature from 27.6 wt % (as ZnSO₄) at 7°C to 41.4 wt % at 39°C. In this range, the heptahydrate is the solid phase. As the temperature rises, the solid phase becomes the hexahydrate and its solubility increases to a maximum of 47.7 wt % at 70°C. Above this temperature, the solid phase is the monohydrate and solubility declines with temperature to 44.0 wt % at the boiling point (105°C).

Many basic zinc sulfates have been reported but probably the only true compounds are hydrates of $3\text{Zn}(\text{OH})_2 \cdot \text{ZnSO}_4$ [12027-98-4] (60).

Zinc sulfate was made by 15 companies in 1980 from secondary materials (93° o) and from roasted ore, ie, zinc oxide (7° o). The zinciferous material reacts with sulfuric acid to form a solution, which is purified. After filtration, the solution is heated to evaporation and heptahydrate crystals are separated. It is sometimes sold in this form but usually as the monohydrate [7446-19-7], which is made by dehydration at ca 100°C. Very pure zinc sulfate solution is made in the manufacture of the pigment lithopone [1345-05-7], ZnSBaSO_4 , and of zinc by electrowinning (see ZINC AND ZINC ALLOYS).

Zinc sulfate is used in fertilizers (qv), sprays, and animal feeds in which it serves as a valuable trace element and disease-control agent. In the manufacture of rayon, it is a crenulating agent in the precipitation bath. It is also the starting material for the manufacture of many zinc chemicals and is used in textile dyeing and printing, flotation (qv) reagents, electrogalvanizing, paper bleaching, and glue. In 1950, rayon accounted for 46° o of zinc sulfate consumption, in 1960 55° o, and in 1966 39° o. Statistics for rayon are not available beyond 1966 but the use in agriculture grew from 39° o in 1966 to 49° o in 1977, to 72° o in 1979, and to 78° o in 1980. The actual tonnage used in agriculture also has increased steadily.

BIBLIOGRAPHY

"Zinc Compounds" in *ECT* 1st ed., Vol. 15, pp. 275–281 by A. P. Thompson and H. W. Schutz, the Eagle-Picher Co.; "Zinc Compounds" in *ECT* 2nd ed., Vol. 22, pp. 604–613 by A. P. Thompson, Eagle-Picher Industries, Inc.; "Zinc Compounds" in *ECT* 3rd ed., Vol. 24, pp. 851–863, by Thomas B. Lloyd, Gulf and Western Natural Resources Group.

1. M. Farnsworth and C. H. Kline, *Zinc Chemicals*, Zinc Institute, Inc., New York, 1973.
2. *Chem. Mark. Rep.* **220**(9), 40 (1981).
3. F. D. Rossini and co-workers, *Circular 500*, U.S. Bureau of Standards, Washington, D.C., 1952.
4. W. Bond, *J. Appl. Phys.* **36**, 1974 (1964).
5. L. Blok and P. L. DeBruyn, *J. Colloid Interface Sci.* **32**, 518 (1970).
6. T. C. Wilder, *Trans. Metall. Soc. AIME* **245**, 1370 (1969).
7. K. K. Kelley, *U.S. Bur. Mines Bull.*, 584 (1960).
8. D. D. Wagman and co-workers, *U.S. Bureau of Standards Technical Note*, U.S. Bureau of Standards, Washington, D.C., 1968, pp. 270–273.
9. P. Pascal, *Nouveau Traite de Chemic Minerals*, Vol. V, Masson et Cie, Paris, 1956.
10. M. Seitz and D. Whitmore, *Phys. Chem. Solids* **29**, 1033 (1968).
11. *Ceram. Ind. (Chicago)* **84**, 157 (1965).
12. J. Turkevich and P. W. Selwood, *J. Am. Chem. Soc.* **63**, 1077 (1941).
13. H. Ibach, *Solid State Commun.* **4**, 353 (1966).
14. W. J. Hamer, *J. Electroanal. Chem.* **10**, 140 (1965).
15. G. F. A. Stutz, *J. Franklin Inst.* **202**, 89 (1926).
16. H. E. Brown, *Zinc Oxide, Properties and Applications*, International Lead Zinc Organization, Inc., New York, 1976, pp. 56–59.
17. *Ibid.*, pp. 88–89.
18. S. B. Salvin, *Ind. Eng. Chem.* **36**, 336 (1944).
19. A. M. van Rij and W. J. Pories in J. O. Nriagu, ed., *Zinc in the Environment*, John Wiley & Sons, Inc., New York, 1980, Part 2, pp. 215–236.
20. E. H. Bunce and H. M. Haslam, *Trans.. AIME* **121**, 678 (1936).
21. C. D. Holley, *The Lead and Zinc Pigments*, John Wiley & Sons, Inc., New York, 1909, pp. 152–153.
22. M. Morton, ed., *Introduction to Rubber Technology*, Reinhold Publishing Co., New York, 1959, p. 222.
23. A. P. Thompson in C. H. Mathewson, ed., *Zinc*, Reinhold Publishing Co., New York, 1959, p. 646.
24. J. H. Calbeck in ref. 23, pp. 344–367.
25. **U.S. Pat. 1,522,097** (Jan. 6, 1925), F. G. Breyer, E. C. Gaskill, and J. A. Singmaster (to The New Jersey Zinc Co.).
26. **U.S. Pat. 959,924** (May 31, 1910), E. Dedolph.
27. W. E. Harris, *Trans. AIME* **121**, 702 (1936).
28. C. W. Morrison in ref. 23, pp. 298–306.
29. **U.S. Pats. 1,743,886** (Jan. 14, 1930); **1,773,779** (Aug. 26, 1930); **1,775,591** (Sept. 9, 1930), E. C. Gaskill (to St. Joseph Lead Co.).
30. J. J. Rankin in ref. 23, pp. 374–384.
31. R. E. Lund and co-workers in C. H. Cotterill and J. M. Cigan, eds., *Extractive Metallurgy of Lead and Zinc*, Vol. II, AIME, New York, 1970, pp. 549–580.
32. **U.S. Pats. 1,522,096; 1,522,098** (Jan. 6, 1925), F. G. Breyer, E. C. Gaskill, and J. A. Singmaster (to The New Jersey Zinc Co.).
33. W. A. Thomas and W. A. Handwerk, *Min. Eng. (N.Y.)* **5**, 1203 (1953).
34. T. R. Janes in ref. 23, p. 385.
35. H. E. Lee and W. T. Isbell in ref. 23, pp. 307–314.
36. **U.S. Pat. 2,303,330** (Dec. 1, 1942), B. R. Silver and E. R. Bridgewater (to The New Jersey Zinc Co. and E. I. du Pont de Nemours & Co., Inc.).
37. **U.S. Pat. 1,997,925** (April 16, 1935), A. C. Eide (to American Zinc Lead and Smelting Co.).
38. **U.S. Pat. 2,251,869** (Aug. 5, 1941), D. L. Gamble and J. H. Haslam (to The New Jersey Zinc Co.).
39. **U.S. Pat. 4,270,955** (June 2, 1981), D. M. Eshelman (to The New Jersey Zinc Co.).
40. **U.S. Pat. 2,372,367** (March 27, 1945), H. A. Depew (to American Lead Zinc & Smelting Co.).
41. *Minerals Yearbook*, U.S. Bureau of Mines, Washington, D.C., 1960, 1965, 1970, 1975, 1980, 1995.
42. *Chem. Market. Rep.* **198**, 43 (Dec. 28, 1970); **208**, 35 (Dec. 29, 1975); **214**, 37 (Dec. 25, 1978); **216**, 37 (Dec. 31, 1979); **218**, 37 (Dec. 29, 1980); **250**, 38 (Dec. 30, 1996).
43. *Zinc—Medical and Biologic Effects of Environmental Pollutants*, Subcommittee on Zinc, National Research Council, University Park Press, Baltimore, Md., 1979, pp. 251–253.
44. P. Drinker and co-workers, *J. Ind. Hyg.* **3**, 98 (1927).
45. R. P. Batchelor and co-workers, *J. Ind. Hyg.* **8**, 322 (1926).
46. W. D. Kelly, *Nat. Saf. News* **119**, 65 (1979).
47. Ref. 16, pp. 6–7.
48. R. G. Silvey and G. E. Myers in ref. 19, Part 2, pp. 237–253.
49. Ref. 19, Parts 1 and 2.
50. Ref. 43, Chapt. 7.
51. G. J. Brewer and A. S. Prasad, eds., *Progress in Clinical and Biological Research*, Vol. 14, A. R. Liss, New York, 1977.
52. Ref. 43, pp. 73–83.
53. Ref. 43, p. 123.
54. Ref. 19, Part 2, pp. 4–5.
55. Ref. 43, p. 187.
56. Ref. 43, p. 249.
57. Ref. 19, Part 2, pp. 45–53.

58. J. W. Hoffman and I. Lander, *Aust. J. Chem.* **21**, 1439 (1968).
59. N. L. Sax, *Dangerous Properties of Industrial Materials*, 5th ed., Van Nostrand Reinhold, New York, 1979, pp. 1100–1104.
60. L. C. Copeland and O. A. Short, *J. Am. Chem. Soc.* **62**, 3285 (1940).

Frank E. Goodwin
International Lead and Zinc Research Organization, Inc.

ZIRCONIUM AND ZIRCONIUM COMPOUNDS

Zirconium [7440-67-7] is classified in subgroup IVB of the periodic table with its sister metallic elements titanium and hafnium. Zirconium forms a very stable oxide. The principal valence state of zirconium is +4, its only stable valence in aqueous solutions. The naturally occurring isotopes are given in Table 1. Zirconium compounds commonly exhibit coordinations of 6, 7, and 8. The aqueous chemistry of zirconium is characterized by the high degree of hydrolysis, the formation of polymeric species, and the multitude of complex ions that can be formed.

Table 1. Naturally Occurring Zirconium Isotopes

Isotope	CAS Registry Number	Occurrence, % ^a	Thermal neutron capture cross section, 10 ⁻²⁸ m ^{2b}
⁹⁰ Zr	[13982-15-5]	51.45	0.03
⁹¹ Zr	[14331-93-2]	11.32	1.14
⁹² Zr	[14392-15-5]	17.19	0.21
⁹⁴ Zr	[14119-12-1]	17.28	0.055
⁹ Zr	[15691-06-2]	2.76	0.020

^a Ref. 1.

^b To convert m² to barns, multiply by 10²⁸.

Zirconium occurs naturally as a silicate in zircon [1490-68-2], the oxide baddeleyite [12036-23-6], and in other oxide compounds. Zircon is an almost ubiquitous mineral, occurring in granular limestone, gneiss, syenite, granite, sandstone, and many other minerals, albeit in small proportion, so that zircon is widely distributed in the earth's crust. The average concentration of zirconium in the earth's crust is estimated at 220 ppm, about the same abundance as barium (250 ppm) and chromium (200 ppm) (2).

Zircon has been known as a gem mineral since biblical times and was known as jargon in Sri Lanka (Ceylon) and as hyacinth in France. The name zircon possibly comes from the Arabic *zayron* for the gold or dark amber color of the more common gemstone. Zircons may be colorless, amber, red, reddish brown, blue, green, or black. In 1789, Klaproth announced that in analyzing jargon from Sri Lanka he had found 68% of an unknown earth which he called zirkonerde (3). In 1797, Vauquelin studied this new earth, to which the name zirconia was given, and published the preparation and properties of some of its compounds (4). In 1824, Berzelius prepared the first crude zirconium metal, a black powder, by heating potassium and potassium hexafluorozirconate [16923-95-8] in a closed pot (5). In 1914, the first relatively pure zirconium was prepared by the reduction of zirconium tetrachloride <casrn>[10026-11-6] with sodium in a bomb.

High purity zirconium was first produced by van Arkel and de Boer in 1925. They vaporized zirconium tetraiodide [13986-26-0] into a bulb containing a hot tungsten filament which caused the tetraiodide to dissociate, depositing zirconium on the filament.

Zirconium is used as a containment material for the uranium oxide fuel pellets in nuclear power reactors (see NUCLEAR REACTORS). Zirconium is particularly useful for this application because of its ready availability, good ductility, resistance to radiation damage, low thermal-neutron absorption cross section 18 × 10⁻³⁰ m² (0.18 barns), and excellent corrosion resistance in pressurized hot water up to 350°C. Zirconium is used as an alloy strengthening agent in aluminum and magnesium, and as the burning component in flash bulbs. It is employed as a corrosion-resistant metal in the chemical process industry, and as pressure-vessel material of construction in the ASME Boiler and Pressure Vessel Codes.

Occurrence and Mining

Zirconium is found in at least 37 different mineral forms (6) but the predominant commercial source is the mineral zircon, zirconium orthosilicate. Other current mineral sources are baddeleyite and eudialyte [12173-26-1].

Zircon occurs worldwide as an accessory mineral in igneous, metamorphic, and sedimentary rocks. Weathering has resulted in segregation and concentration of the heavy mineral sands in layers or lenses of placer deposits in river beds and ocean beaches. All commercial sources of zircon are derived from the mining of these ancient, unconsolidated beach deposits, the largest of which are in Kerala State in India, Sri Lanka, the east and west Coasts of Australia, on the Trail Ridge in Florida, and at Richards Bay in the Republic of South Africa. These heavy mineral sands are processed for the recovery of the titanium-bearing minerals ilmenite, rutile, and leucoxene, the zircon is obtained as a co-product. The output of zircon depends largely then on the market for these titanium minerals used in producing titanium oxide white pigment and titanium metal.

The deposits, usually ca 4% heavy minerals, are mined with front-end loaders or sand dredges. Typically, the overburden is bulldozed away, the excavation is flooded, and the raw sand is handled by a floating sand dredge capable of dredging to a depth of 18 m. The material is broken up by a cutter head to the bottom of the deposit and the sand slurry is pumped to a wet-mill concentrator mounted on a floating barge behind the dredge. Initial wet concentration using screens, Reichert cones, spirals, and cyclones removes the coarse sand, slimes, and light-density sands to produce a 40 wt % heavy-mineral concentrate. The tailings are returned to the back end of the excavation and used for rehabilitation of worked-out areas. The concentrate is dried and iron oxide and other surface coatings are removed; various combinations of gravity separation, magnetic separation (qv), and electrostatic separation yield individual concentrates of rutile, ilmenite, leucoxene, zircon, monazite, and xenotime. Other heavy minerals such as staurolite, tourmaline, sillimanite, corundum, and magnetite may be recovered as local situations warrant. Typical analyses of zircon sand samples are given in Table 2.

Baddeleyite, a naturally occurring zirconium oxide, has been found in the Poco de Caldas region of the states of Sao Paulo and Minas Geraes in Brazil, the Kola Peninsula of the former USSR, and the northeastern Transvaal of the Republic of South Africa. Brazilian baddeleyite occurs frequently with zircon, and ore shipments are reported to contain 65–85% zirconium oxide, 12–18% silica, and 0.5% uranium oxide. Very little of this ore is exported now because all radioactive minerals are under close control of the Brazilian government.

The Phalaborwa complex in the northeastern Transvaal is a complex volcanic orebody. Different sections are mined to recover magnetite, apatite, a copper concentrate, vermiculite, and baddeleyite, listed in order of annual quantities mined. The baddeleyite is contained in the foskorite ore zone at a zirconium oxide concentration of 0.2%, and at a lesser concentration in the carbonatite orebody. Although baddeleyite is recovered from the process tailings to meet market demand, the maximum output could be limited by the requirements for the magnetite and apatite. The baddeleyite concentrate contains ca 96% zirconium oxide with a hafnium content of 2% Hf/Zr + Hf. A comminuted, chemically beneficiated concentrate containing ca 99% zirconium oxide is produced also.

Table 2. Typical Analysis of Zircon Sands, %

Assay	Sri Lanka	India	Nigeria	Republic of South Africa	United States, Florida	Australia	
						East	West
constituents (Zr + Hf)O ₂	64.9	64.4	58.2	65.0	65.3	65.7	64.6
Fe ₂ O ₃	0.16	0.23	0.90	0.18	0.07	0.06	0.28
Al ₂ O ₃	0.20	0.60	0.65	0.15	0.17	0.14	0.25
TiO ₂	0.71	0.16	0.05	0.08	0.10	0.17	0.32
P ₂ O ₅	0.18	0.13	0.39	0.13	0.05	0.07	0.08
U ₃ O ₈	0.03	0.04	0.11	0.03	0.03	0.03	0.03
Nb ₂ O ₅	<0.01	<0.01	2.0	<0.01	<0.01	<0.01	<0.01
Hf Hf + Zr	2.2	2.3	6.8	2.3	2.2	2.2	2.2
size, μm (mesh)							
>149 (+100)	0.1	41.8	97.2	5.6	0.4	83.8	20.2
<74–149 (–100–200)	40.7	48.6	2.7	92.7	91.7	14.2	77.9
<74 (–200)	59.2	9.6	0.1	1.7	7.9	2.0	1.9

Eudialyte, (Na,Ca) ZrOH(Si₅O₉)₂, from a large deposit near Narssaq in southwest Greenland, is the source of pure zirconium oxide. The hafnium ratio in the ore is 2.2^o o Hf Zr + Hf.

Two zirconium-containing minerals were discovered in North America, namely, welognite [55659-01-3], (Sr_{2.8}Ca_{0.2})ZrNa₂(CO₃)₆·3H₂O, and gittinsite [75331-27-0], CaZrSi₂O₇ (7,8). Unlike zircon, the zirconium content of these minerals and eudialyte can be dissolved by strong acid.

Physical Properties

Zirconium is a hard, shiny, ductile metal, similar to stainless steel in appearance. It can be hot-worked to form slabs, rods, and rounds from arc-melted ingot. Further cold-working of zirconium with intermediate annealings produces sheet, foil, bar wire, and tubing. Physical properties are given in Table 3.

Chemical Properties

Zirconium forms anhydrous compounds in which its valence may be 1, 2, 3, or 4, but the chemistry of zirconium is characterized by the difficulty of reduction to oxidation states less than four. In aqueous systems, zirconium is always quadrivalent. It has high coordination numbers, and exhibits hydrolysis which is slow to come to equilibrium, and as a consequence zirconium compounds in aqueous systems are polymerized.

Zirconium is a highly active metal which, like aluminum, seems quite passive because of its stable, cohesive, protective oxide film which is always present in air or water. Massive zirconium does not burn in air, but oxidizes rapidly above 600°C in air. Clean zirconium plate ignites spontaneously in oxygen of ca 2 MPa (300 psi); the autoignition pressure drops as the metal thickness decreases. Zirconium powder ignites quite easily. Powder (<44 μm or –325 mesh) prepared in an inert atmosphere by the hydride–dehydride process ignites spontaneously upon contact with air unless its surface has been conditioned, ie, preoxidized by slow addition of air to the inert atmosphere. Heated zirconium is readily oxidized by carbon dioxide, sulfur dioxide, or water vapor.

Table 3. Physical Properties of Zirconium

Property	Value	Refs.
atomic weight	91.22	
density at 298.15 K, g cm ³	6.5107	3,5
crystal structure		
αZr		9
close-packed hexagonal space	P6 ₃ mmc	10
group		
a, nm	0.3231	
c, nm	0.5146	
c a	1.5927	
βZr		11
body-centered cubic space group	Im3m	9
α-β transition temperature, K	1136 ± 5	12
melting temperature, K	2125 ± 10	13
boiling temperature, K	4577 ± 100	13
vapor pressure, T < 2125 K,		14
log ₁₀ P _{kPa} ^a		
βZr	8.956 ± 0.080-(30810 ± 240) T ⁻¹	
liquid Zr	8.547 ± 0.080-(29940 ± 240) T ⁻¹	
heat of transition, kJ mol ^b	3.89 ± 0.08	13
heat of melting, kJ mol ^b	18.8 ± 2.1	14
heat of boiling, kJ mol ^b	573.2 ± 4.6	15
heat of sublimation at 298 K, kJ mol ^b	600.8	15
heat capacity, T = 298–1136, J (mol·K) ^b		12
αZr	22.857 ± 8.970 × 10 ⁻³ T – 0.69 × 10 ⁵ T ⁻²	
βZr	21.493 ± 6.586 × 10 ⁻³ T + 36.718 × 10 ⁵ T ⁻²	
entropy at 298.15 K, J (mol·K) ^b	38.99 ± 0.46	12
thermal expansion		16
single crystal		
perpendicular to c-axis, L _{TC} L _{0°C}	1 + 5.145 × 10 ⁻⁶ T	
parallel to c-axis, L _{TC} L _{0°C}	1 + 9.213 × 10 ⁻⁶ T – 6.385 × 10 ⁻⁹ T ² + 18.491 × 10 ⁻¹² T ³ – 9.856 × 10 ⁻¹⁵ T ⁴	

volumetric, $V_{TC} = V_{0^\circ C}$	$1 + 19.756 \times 10^{-6} T - 7.023 \times 10^{-9} T^2 + 19.146 \times 10^{-12} T^3$ $9.980 \times 10^{-15} T^4$	
polycrystalline linear, for random orientation, $L_{TC} = L_{0^\circ C}$	$1 + 6.499 \times 10^{-6} T - 2.096 \times 10^{-9} T^2 + 6.108 \times 10^{-12} T^3$ $3.259 \times 10^{-15} T^4$	
thermal conductivity, W (m·K)		17 ^c
0–10 K	$1.08189 \times 10^{-1} T + 1.65524 \times 10^{-3} T^2 - 2.63239 \times 10^{-4} T^3$	
10–25 K	$9.48650 \times 10^{-1} + 3.44316 \times 10^{-1} T - 1.79663 \times 10^{-2} T^2 +$ $2.82821 \times 10^{-4} T^3$	
25–80 K	$1.80754 - 5.50950 \times 10^{-2} T - 7.61839 \times 10^{-4} T^2 - 3.72006 \times$ $10^{-6} T^3$	
80–500 K	$5.18009 \times 10^{-1} - 2.36738 \times 10^{-3} T + 6.28905 \times 10^{-6} T^2$ $5.58159 \times 10^{-9} T^3$	
500–1900 K	$2.44486 \times 10^{-1} - 2.3982 \times 10^{-4} T + 2.27218 \times 10^{-7} T^2$ $6.24923 \times 10^{-11} T^3$	
electrical resistivity, Zr, at 5°C, Ω·cm	$43.740.08 \times 10^{-6}$	18
temperature coefficient, 0–200°C, per °C	42.5×10^{-4}	19
elastic moduli, Zr, GPa ^d		20–21
single-crystal adiabatic, 19.7°C, 6.505 g cm ³		
C_{11}	143.5 ± 0.2	
C_{12}	72.5 ± 0.2	
C_{13}	65.4 ± 0.2	
C_{33}	164.9 ± 0.2	
C_{44}	32.07 ± 0.03	
polycrystalline adiabatic ^e		20
Young's modulus, GPa ^e	97.1	
shear modulus, GPa ^e	36.5	
bulk modulus, GPa ^e	954	
Poisson's ratio	0.33	22
Brinell hardness number, HB ^f	90–130	

^a To convert kPa to mm Hg, multiply by 7.5.

^b To convert J to cal, divide by 4.184.

^c Tabulated selected experimental values were regressed to give the equation quoted. Estimated values tabulated were not used in the regression. All equations represent tabulated, nonestimated values within ±2% or better.

^d To convert GPa to psi, multiply by 145,000.

^e Calculated from single-crystal adiabatic moduli by the Voight method.

^f At good Kroll-process purity, lower for iodide zirconium.

Zirconium reacts more slowly with nitrogen than with oxygen. Heating in nitrogen for 3 min gives a 0.3 μm layer of zirconium nitride [25658-42-8] at 700°C or a 1.2-μm layer at 900°C. The nitriding rate is enhanced by the presence of oxygen in the nitrogen or on the metal surface. Clean zirconium in ultrapure nitrogen reacts more slowly. Although the nitride reaction occurs at 900°C or higher, diffusion of nitrogen into zirconium is slow, and temperatures of 1300°C are needed to fully nitride the metal.

Heated zirconium is readily chlorinated by ammonium chloride, molten stannous chloride, zinc chloride, and chlorinated hydrocarbons and the common chlorinating agents. It is slowly attacked by molten magnesium chloride in the absence of free magnesium, which is always present in the Kroll process.

Zirconium is readily attacked by acidic solutions containing fluorides. As little as 3 ppm fluoride ion in 50% boiling sulfuric acid corrodes zirconium at 1.25 mm yr. Solutions of ammonium hydrogen fluoride or potassium hydrogen fluoride have been used for pickling and electropolishing zirconium. Commercial pickling is conducted with nitric–hydrofluoric acid mixtures (see METAL SURFACE TREATMENTS).

Corrosion Resistance. Zirconium is resistant to corrosion by water and steam, mineral acids, strong alkalis, organic acids, salt solutions, and molten salts (28) (see also CORROSION AND CORROSION CONTROL). This property is attributed to the presence of a dense adherent oxide film which forms at ambient temperatures. Any break in the film reforms instantly and spontaneously in most environments.

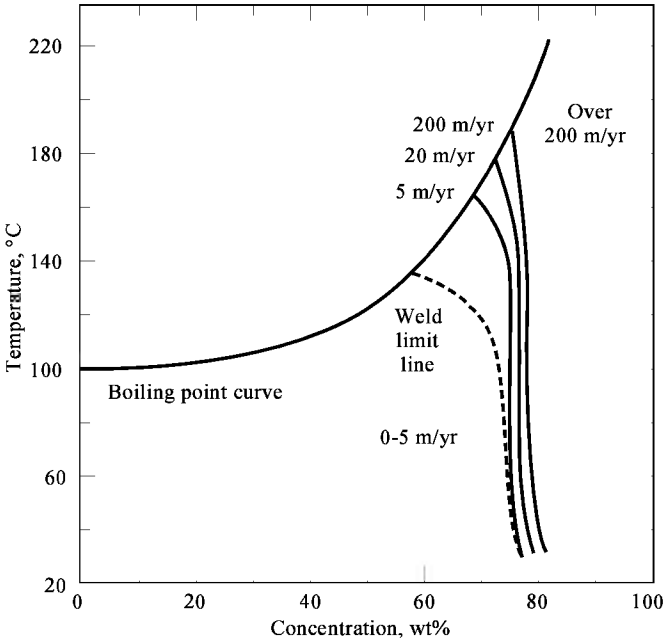


Fig. 1. Corrosion of zirconium in sulfuric acid.

Zirconium is completely resistant to sulfuric acid up to boiling temperatures, at concentrations up to 70 wt % , except that the heat-affected zones at welds have lower resistance in >55 wt % concentration acid (Fig. 1). Fluoride ions must be excluded from the sulfuric acid. Cupric, ferric, or nitrate ions significantly increase the corrosion rate of zirconium in 65–75 wt % sulfuric acid.

Zirconium resists attack by nitric acid at concentrations up to 70 wt % and up to 250°C. Above concentrations of 70 wt % , zirconium is susceptible to stress-corrosion cracking in welds and points of high sustained tensile stress (29). Otherwise, zirconium is resistant to nitric acid concentrations of 70–98 wt % up to the boiling point.

Zirconium is not attacked by caustics up to boiling temperatures. It is resistant to molten sodium hydroxide to 1000°C, but is less resistant to potassium hydroxide.

Zirconium is totally resistant to corrosion by organic acids. It has been used in urea-production plants for more than two decades.

Zirconium is totally resistant to attack of hydrochloric acid in all concentrations to temperatures well above boiling (Fig. 2). Aeration has no effect, but oxidizing agents such as cupric or ferric ions may cause pitting. Zirconium also has excellent corrosion resistance to hydrobromic and hydriodic acid.

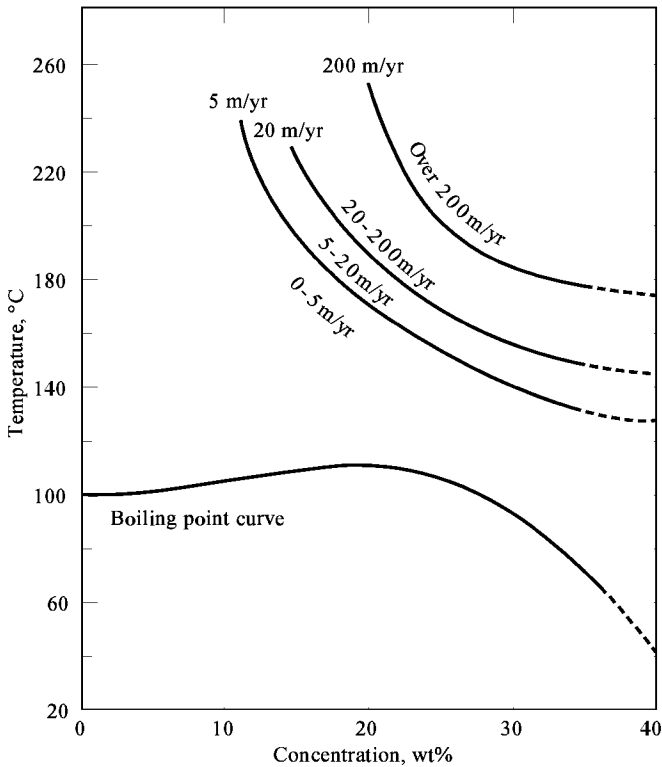
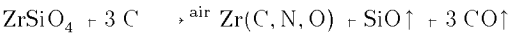


Fig. 2. Corrosion of zirconium in hydrochloric acid.

Processing

Decomposition of Zircon. Zircon is a highly refractory mineral as shown by its geological stability; the ore is cracked only with strong reagents and high temperature.

Electric Furnace. Zircon and coke have reacted in an electric arc furnace to produce a crude zirconium carbide nitride [12713-24-5] (ca 6 wt % C, 2 wt % N, 1 wt % O):



With a deficiency of carbon, the silica is not reduced to carbide but converted into silicon monoxide which is vaporized at the reaction temperature, estimated to be 2500°C. In actual operation, the above proportions are optimal. The additional carbon is derived from the graphite electrodes. If more coke is added, more silicon is retained as the carbide in the fused ingot. When feed of the batch charge is completed, the fused ingot is separated from the unreacted charge which had insulated the ingot from the furnace shell, and the hot ingot is allowed to oxidize to crude zirconia. Alternatively, the entire charge is cooled and separated and the ingot is broken into lumps which are subsequently chlorinated.

Mixed zircon, coke, iron oxide, and lime reduced together produce zirconium ferrosilicon [71503-20-3], 15 wt % Zr, which is an alloy agent. Fused zirconia [1314-23-4] has been made from zircon but baddeleyite is now the preferred feed for the production of fused zirconia and fused alumina–zirconia by electric-arc-furnace processing.

Caustic Fusion. Fusion of finely ground zircon with caustic soda at 600°C or with soda ash produces a frit containing sodium silicate, sodium zirconate [12201-48-8], and some sodium silicozirconate (25,26). Water removes most of the sodium and silica, leaving a hydrous zirconium oxide which is soluble in most mineral acids. If the fusion is conducted with less alkali, the resulting frit is essentially Na₂ZrSiO₅ [12027-83-7] which can be ground and treated with a strong mineral acid to solubilize and extract the zirconium. These two methods are most commonly used to produce aqueous zirconium solutions, hydrated zirconium compounds, and zirconium oxide.

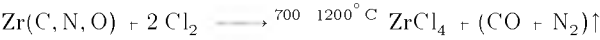
Similarly, fusion of milled zircon with dolomite or lime forms CaSiO₃ and MgZrO₃ [12032-31-4], CaZrO₃ [12013-47-7], and CaO·Ca₂SiO₄ or CaSiO₃ and ZrO₂, and is used to prepare zirconium oxide, usually as calcia-stabilized cubic zirconia because of the calcia left in solid solution in the zirconia (27–29).

Fluorosilicate Fusion. The fusion reaction of milled zircon with potassium hydrogen fluoride was used to prepare potassium hexafluorozirconate [16923-95-8] for studies leading to the first separation of hafnium and zirconium (30). Similar reactions using potassium hexafluorosilicate have been used (31,32) commercially in the United States and the former USSR:

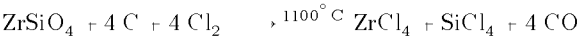


The use of potassium hexafluorosilicate is preferred over sodium hexafluorosilicate because of the lower tendency of the potassium compound to dissociate the lose silicon tetrafluoride by sublimation. The addition of potassium carbonate or chloride to the fusion mix further reduces this tendency and promotes completion of the reaction. The reaction is conducted in a rotary furnace operating at 700°C. The product is crushed prior to leaching with acidified hot water. The hot slurry is filtered to remove the silica, and potassium hexafluorozirconate crystallizes as the solution cools.

Chlorination. Historically, the production of zirconium tetrachloride from zircon sand involved first a reduction to carbide nitride (see above) followed by the very exothermic reaction of the crushed carbide nitride with chlorine gas in a water-cooled vertical shaft furnace:



In the current practice, milled zircon and coke are chlorinated in fluidized beds using chlorine as the fluidizing medium:



Additional energy to sustain the endothermic reaction is provided chemically by the addition of silicon carbide grain or electrically by use of electrothermal fluidized beds (33–34), induction heating, or resistance heating. Chlorine efficiencies are typically 98% or better.

The product gases are first cooled below 200°C to selectively condense so-called zirconium tetrachloride snow in a large space condenser. The silicon tetrachloride subsequently is condensed in a quench condenser wherein the warm gases are countercurrently scrubbed with liquid silicon tetrachloride at –20°C. The silicon tetrachloride is purified by stripping and distillation.

Thermal Dissociation. The thermal dissociation of zircon into zirconia and cristobalite or liquid silica above 1650°C has been studied as a means of producing zirconia, but the process was hampered by the partial recombination of the two phases during cooling. Passing particulate zircon through the intense heat of a plasma and immediately thereafter quenching the molten droplet results in the formation of dissociated zircon particles of generally spheroidal configuration (35). The spheroids consist of a fused, amorphous silica mass in which radiating crystallites of monoclinic zirconia are dispersed.

This dissociated zircon is amenable to hot aqueous caustic leaching to remove the silica in the form of soluble sodium silicate. The remaining skeletal structure of zirconia is readily washed to remove residual caustic. Purity of this zirconia is directly related to the purity of the starting zircon since only silica, phosphate, and trace alkalies and alkaline earth are removed during the leach. This zirconia, and the untreated dissociated zircon, are both proposed for use in ceramic color glazes (36) (see COLORANTS FOR CERAMICS).

Separation of Hafnium. Zirconium and hafnium always occur together in natural minerals and therefore all zirconium compounds contain hafnium, usually about 2 wt % Hf/Hf + Zr. However, the only applications that require hafnium-free material are zirconium components of water-cooled nuclear reactors.

Zirconium and hafnium have very similar chemical properties, exhibit the same valences, and have similar ionic radii, ie, 0.074 nm for Zr⁴⁺, 0.075 nm for Hf⁴⁺ (see HAFNIUM AND HAFNIUM COMPOUNDS). Because of these similarities, their separation was difficult (37–40). Today, the separation of zirconium and hafnium by multistage counter-current liquid–liquid extraction is routine (41) (see EXTRACTION, LIQUID–LIQUID).

In the initial thiocyanate-complex liquid–liquid extraction process (42,43), the thiocyanate complexes of hafnium and zirconium were extracted with ether from a dilute sulfuric acid solution of zirconium and hafnium to obtain hafnium. This process was modified in 1949–1950 by an Oak Ridge team and is still used in the United States. A solution of thiocyanic acid in methyl isobutyl ketone (MIBK) is used to extract hafnium preferentially from a concentrated zirconium–hafnium oxide chloride solution which also contains thiocyanic acid. The separated metals are recovered by precipitation as basic zirconium sulfate and hydrous hafnium oxide, respectively, and calcined to the oxide (44,45). This process is used by Teledyne Wah Chang Albany Corporation and Western Zirconium Division of Westinghouse, and was used by Carborundum Metals Company, Reactive Metals Inc., AMAX Specialty Metals, Toyo Zirconium in Japan, and Pechiney Ugine Kuhlmann in France.

In the tributyl phosphate extraction process developed at the Ames Laboratory, Iowa State University (46–48), a solution of tributyl phosphate (TBP) in heptane is used to extract zirconium preferentially from an acid solution (mixed hydrochloric–nitric or nitric acid) of zirconium and hafnium (45). Most other impurity elements remain with the hafnium in the aqueous acid layer. Zirconium recovered from the organic phase can be precipitated by neutralization without need for further purification.

High molecular weight primary, secondary, and tertiary amines can be employed as extractants for zirconium and hafnium in hydrochloric acid (49–51). With similar aqueous-phase conditions, the selectivity is in the order tertiary > secondary > primary amines. The addition of small amounts of nitric acid increases the separation of zirconium and hafnium but decreases the zirconium yield. Good extraction of zirconium and hafnium from ca 1 M sulfuric acid has been effected with tertiary amines (52–54), with separation factors of 10 or more. A system of this type, using trioctylamine in kerosene as the organic solvent, is used by Nippon Mining of Japan in the production of zirconium (55).

Zirconium and hafnium are separated by fractional distillation of the anhydrous tetrachlorides in a continuous molten solvent salt KCl–AlCl₃ system at atmospheric pressure (56,57). Zirconium and hafnium tetrachlorides are soluble in KCl–AlCl₃ without compound formation and are produced simultaneously.

Pure zirconium tetrachloride is obtained by the fractional distillation of the anhydrous tetrachlorides in a high pressure system (58). Commercial operation of the fractional distillation process in a batch mode was proposed by Ishizuka Research Institute (59). The mixed tetrachlorides are heated above 437°C, the triple point of zirconium tetrachloride. All of the hafnium tetrachloride and some of the zirconium tetrachloride are distilled, leaving pure zirconium tetrachloride. The innovative aspect of this operation is the use of a double-shell reactor. The autogenous pressure of 3–4.5 MPa (30–45 atm) inside the heated reactor is balanced by the nitrogen pressure contained in the cold outer reactor (60). However, previous evaluation in the former USSR of the binary distillation process (61) has cast doubt on the feasibility of also producing zirconium-free hafnium tetrachloride by this method because of the limited range of operating temperature imposed by the small difference in temperature between the triple point, 433°C, and critical temperature, 453°C, a hafnium tetrachloride.

Reduction. Brezelius attempted the first reduction of zirconium in 1824 by the reaction of sodium with potassium fluorozeonate. However, the first pure ductile metal was made in 1925 by the iodide thermal-dissociation method. The successful commercial production of pure ductile zirconium via the magnesium reduction of zirconium tetrachloride vapor in an inert gas atmosphere was the result of the intense research efforts of Kroll and co-workers at the U.S. Bureau of Mines in 1945–1950 (62–67).

Obtaining pure ductile zirconium by reduction of the oxide is particularly difficult because of the tendency of hot zirconium to dissolve considerable amounts of oxygen, making the metal brittle at room temperature. Therefore, it is common practice to reduce oxygen-free zirconium tetrachloride.

Kroll Process. Hafnium-free zirconium dioxide is mixed with pulverized coke and fed into an induction-heated chlorinator where the mixture is fluidized by chlorine gas; reaction at 900°C gives zirconium tetrachloride and carbon dioxide. The product gas passes through a nickel-lined condenser where zirconium tetrachloride powder is formed by cooling below 200°C. It is purified by subliming and recondensing in a nitrogen–hydrogen atmosphere to reduce the aluminum and phosphorus contents. The tetrachloride powder is charged into the upper chamber of a vertical cylindrical steel retort which contains cast ingots of magnesium in a stainless-steel liner within the lower chamber. The retort is sealed and evacuated and backfilled with argon several times at 200°C. Heat applied to the lower retort chamber melts the magnesium which in turn reduces the zirconium tetrachloride vapors as they sublime out of the upper chamber. The retort is cooled and unloaded. The stainless-steel liner is peeled from the reduction mass and the bulk of the magnesium chloride salt is physically separated from the Zr–Mg regulus. Several reduction reguli are stacked and loaded into a furnace for removal of residual magnesium chloride. The distillation is conducted in a vacuum <1.3 Pa (<10 μm Hg). As the temperature is gradually increased to 980°C, the magnesium chloride melts and drains and the magnesium metal is distilled and condensed on the cold lower retort wall, leaving porous zirconium sponge. After cooling and conditioning, the sponge is removed from the retort and broken into chunks using a hydraulic chisel. The chunks are graded, crushed to ca 1 cm dia, and sampled before blending and melting.

Other Reductions. Ductile, pure zirconium has been made by a two-stage sodium reduction of zirconium tetrachloride (68) in which the tetrachloride and sodium are continuously fed into a stirred reactor to form zirconium dichloride [13762-26-0]; heating with additional sodium yields zirconium metal. Leaching with water removes the sodium chloride from the zirconium. Bomb reduction of pure zirconium tetrafluoride with calcium also produces pure metal (69).

Finely divided zirconium powder is made by bomb reduction of zirconium oxide with magnesium or calcium. The powder is separated by leaching

with cold hydrochloric acid. Because of its large surface area, it is extremely pyrophoric and hazardous to produce. Zirconium powder made by this technique is used in ordnance fuses and incendiaries, but it contains considerable oxygen and therefore cannot be converted into ductile mill products.

Electrolysis. Electrowinning of zirconium has long been considered as an alternative to the Kroll process, and at one time zirconium was produced electrolytically in a prototype production cell (70). Electrolysis of an all-chloride molten-salt system is inefficient because of the stability of lower chlorides in these melts. The presence of fluoride salts in the melt increases the stability of Zr^{4+} in solution, decreasing the concentration of lower valence zirconium ions, and results in much higher current efficiencies. The chloride–electrolyte systems and electrolysis approaches are reviewed in References 71 and 72. The recovery of zirconium metal by electrolysis of aqueous solutions is not thermodynamically feasible, although efforts in this direction persist.

Refining. Zirconium sponge produced by the Kroll process has adequate purity and ductility for most uses. For applications requiring extremely soft metal and for research studies on the properties of the pure metal, it can be further purified by the van Arkel-de Boer (iodide-bar) process using a selective vapor transport (73–76). Zirconium sponge is loaded into a cylindrical Inconel vessel. The vessel lid contains insulated electrical lines from which a zirconium wire filament in a hairpin shape is suspended. Iodine is added to the evacuated vessel which then is heated to 250°C in a molten-salt bath. Volatile zirconium tetraiodide forms and diffuses to the central zirconium filament which is resistance-heated to 1200–1500°C. The tetraiodide thermally dissociates at the hot filament, depositing zirconium and releasing iodine to react again with the sponge feed. The deposition rate is controlled by the feed-bed temperature, filament temperature, iodine concentration, filament-to-bed distance, and the presence of other gases. Ordinarily, bars of 40 mm dia are grown from 3-mm filament wire. Under conditions optimized for zirconium transfer, the impurity metals and metallic oxides, carbides, and nitrides transfer poorly so that the zirconium filament is much softer and purer than the starting sponge.

Electron-beam melting of zirconium has been used to remove the more volatile impurities such as iron, but the relatively high volatility of zirconium precludes effective purification. Electrorefining is fused-salt baths (77,78) and purification by d-c electrotransport (79) have been demonstrated but are not in commercial use.

Economic Aspects

Normally, zircon sand is readily available as a by-product of rutile and ilmenite mining at ca \$150 per metric ton. However, zircon and baddeleyite are obtained as by-products of their operations, and therefore, the supply is limited by the demand for other minerals. In 1974, when a use for zircon in tundish nozzles developed in the Japanese steel industry, a resulting surge in demand and stockpiling raised zircon prices to \$500 /t. Worldwide production by country is given in Reference 80.

Specifications and Standards

The U.S. specifications for zirconium are listed in Table 4. For nuclear power use, each reactor vendor issues particular, detailed specifications which usually include the pertinent ASTM nuclear specifications.

Table 4. U.S. Specifications for Zirconium

Form	ASTM		ASME	DOE	AWS ^a
	Nuclear	Commercial			
sponge	B 349	B 494			
ingot	B 350	B 495		M 10-1T	
bars, rod, and wire	B 351	B 550	SB550	M 7-9T	
flat rolled products	B 352	B 551	SB551	M 5-6T	
tubing	B 353	B 523	SB523	M 3-8T	
forging and extrusions		B 493	SB493	M 2-9T	
bare welding rods			SFA5.24	M 1-16T	A5.24
descaling and cleaning		B 614			
welding fittings		B 653			
seamless and welded pipe		B 658	SB658		
aqueous corrosion testing	G 2	G 2			

^a American Welding Society.

The four ASTM-specified grades of commercial zirconium and zirconium alloys are given in Table 5. The unalloyed Zr 702, the most commonly used commercial grade, has the best overall chemical corrosion resistance but it is the lowest in strength; Zr 705 has similar corrosion resistance in most environments but its strength is almost double that of Zr 702. It has better formability in applications requiring bending through a sharp radius. Both Zr 702 and 705 are approved for use in the construction of pressure vessels according to the American Society of Mechanical Engineers Boiler and Pressure Vessel Code, Section VIII.

Table 5. ASTM Grades and Commercial Designations of Commercial Zirconium and Alloys

Nominal composition, wt %	R 60702, Zr 702	R 60704, Zr 704	R 60705, Zr 705	R 60706, Zr 706
Zr + Hf, min	99.2	97.5	95.2	95.5
Fe + Cr	<0.2	0.3	<0.2	<0.2
Sn		1.5		
Nb			2.5	2.5
Hf	2.0	2.0	2.0	2.0

The chemical corrosion resistance of Zr 704 is slightly less than that of Zr 702 in some environments but Zr 704 is superior in high temperature, high pressure water, and steam. A softer version of Zr 705 is Zr 706, developed specifically for severe forming applications such as panel-type heat exchangers.

There are no industry-wide specifications for zirconium metal castings, or for zirconium chemicals.

For nuclear applications, hafnium-free zirconium is used, mostly as Zircaloy 2 and Zircaloy 4, in alloys developed in the U.S. Naval Nuclear Propulsion Program. The nickel-free Zircaloy 4 absorbs less of the hydrogen generated by steam corrosion, and is used increasingly. Zr-2.5Nb was adapted by Atomic Energy of Canada, Ltd., for use in Candu reactor pressure tubes (see NUCLEAR REACTORS). Excel is a newer, stronger, creep-resistant alloy developed in Canada as a pressure-tube material. Ozhennite 0.5 is another zirconium alloy from the former USSR for nuclear applications. A UK zirconium alloy, A.T.R., is used in carbon dioxide–cooled reactors. The alloying compositions are given in Table 6.

Table 6. Hafnium-Free Zirconium Alloys for Nuclear Service

Element	Zircaloy 2	Zircaloy 4	Zr-2.5 Nb	Excel	A.T.R.	Ozhennite 0.5
Sn	1.5	1.5	<0.02	3.5	<0.01	0.02
Fe	0.14	0.22	<0.08	<0.08	<0.05	0.1
Cr	0.1	0.1	<0.02	<0.02	<0.01	<0.02
Ni	0.05	<0.004	<0.007	<0.007	<0.004	0.1
Nb			2.5	0.8		0.1
Cu					0.55	
Mo				0.8	0.55	

Analytical Methods

Zirconium is often determined gravimetrically. The most common procedure utilizes mandelic acid (81) which is fairly specific for zirconium plus hafnium. Other precipitants, including nine inorganic and 42 organic reagents, are listed in Reference 82. Volumetric procedures for zirconium, which also include hafnium as zirconium, are limited to either EDTA titrations (83) or indirect procedures (84). X-ray fluorescence spectroscopy gives quantitative results for zirconium, without including hafnium, for concentrations from 0.1 to 50% (85). Atomic absorption determines zirconium in aluminum in the presence of hafnium at concentrations of 0.1–3% (86).

Emission spectroscopy is used for lower concentrations and trace levels. Methods, as outlined in ASTM procedures (87), include zirconium in aluminum and aluminum alloys, ceramics, sand, magnesium alloys, and titanium.

Colorimetric methods always include hafnium. Most methods employ a separation step such as solvent extraction. The three reagents used successfully are 8-hydroxyquinoline (88), alizarin red S (89), and catechol violet (90).

Impurities in zirconium and zirconium alloys and compounds are often determined by emission spectroscopy. Both carrier distillation techniques and point-to-plane methods are available (91,92). Several metallic impurities can be determined instantaneously by this method. Atomic absorption analysis has been used for iron, chromium, tin, copper, nickel, and magnesium (93). The interstitial gases, hydrogen, nitrogen, and oxygen are most often determined by chromatography (81). Procedures for carbon, chloride, fluoride, phosphorus, silicon, sulfur, titanium, and uranium in zirconium are given in the literature (81,94–96).

Health and Safety Factors

Zirconium is generally nontoxic as an element or in compounds (97,98). At pH normally associated with biological activity, zirconium chiefly exists as the dioxide which is insoluble in water and in this form zirconium is physiologically inert.

Chelated complexes such as sodium zirconium lactate [15529-67-6] or ammonium zirconium carbonate [22829-17-0], and acidic forms such as zirconium hydroxy oxide chloride [18428-88-1] have been used in preparations in deodorants or for treatment for poison oak and poison ivy dermatitis. In such occasions, when the skin had been cut or abraded, a few users developed granulomas which have been identified as a delayed hypersensitivity to zirconium (99). These may take several weeks to develop, and commonly persist for 6 months to over a year.

The oral toxicity is low; OSHA standards for pulmonary exposure specify a TLV of 5 mg zirconium per m³.

In massive form, zirconium is commonly brought to red heat before forging, swaging, or other hot working with only slight surface oxidation of the ingot or slab. Conversely, zirconium metal with a high surface area such as sponge, foil, fine powder, fine machining chips, and grinding dust is extremely flammable and often pyrophoric, especially when slightly damp. Many unexpected ignitions have occurred caused by improper handling of finely divided forms of zirconium metal, resulting in more than one incident of fatal flash burns to metal workers. Very fine dusts of zirconium ignite when dispersed in air. Fires can be extinguished with a blanket of argon or a layer of dry salt or sand. Water is hazardous unless the zirconium can be quenched below its ignition temperature. A settled 25-kg mass of 20-µm particles ignited in the bottom of a water-filled 55-gal (208-L) drum burns for hours. A stream of water results in the generation of hydrogen and steam which disperse the burning fragments and spread the fire (100). Safe handling of zirconium has been discussed in several publications (100–102).

Zirconium powder reacts exothermically with many other elements, including hydrogen, boron, carbon, nitrogen, and the halogens, although the ignition temperature is usually above 200°C. The reaction between zirconium powder and platinum is especially violent.

Finely divided zirconium is classified as a flammable solid and shipping regulations are prescribed accordingly (103). Metal powder finer than 74 µm (270 mesh) is limited to 2.26 kg per individual container.

Uses

The largest use is foundry sand, where zircon is used as the basic mold material, as facing material on mold cores, and in ram mixes. Because of its chemical inertness and high melting point, zircon is wetted less easily by molten metal, producing smoother surfaces on iron, high alloy steel, aluminum, and bronze castings. Zircon-sand molds have greater thermal shock resistance and better dimensional stability than quartz-sand molds. The zircon grains are usually bonded with sodium silicate in the formation of the molds and cores, but other binders, eg, clay, lignin-containing liquors, and furan resin, are used. The consumption of zircon in foundry applications was significantly reduced by the 1974 price increases. Alternatives include olivine, chromite, and quartz as well as increased cleaning and recycling of zircon foundry sand.

Fired zircon and AZS (alumina–zirconia–silica) refractories, including brick, crucibles, and saggars, have their main use in the production of handling of molten glasses (see REFRACTORIES). Blends of zircon with alumina or magnesia improve spalling resistance, reduce thermal conductivity, withstand thermal cycling, and have better corrosion resistance in specific glass-furnace environments. Other uses for zircon refractories include hearths and nozzles for handling molten metal and in extrusion dies. Finely milled zircon is used as an opacifier in glazes for tiles and sanitary wares. The use of zircon in these applications is ca 60% of the U.S. consumption of zircon and baddeleyite; the remaining 40% are used in the conversion into other forms of zirconium.

Zirconium oxide is fused with alumina in electric-arc furnaces to make alumina–zirconia abrasive grains for use in grinding wheels, coated-abrasive disks, and belts (104) (see ABRASIVES). The addition of zirconia improves the shock resistance of brittle alumina and toughens the abrasive. Most of the baddeleyite imported is used for this application, as is zirconia produced by burning zirconium carbide nitride.

Zirconium oxide is used in the production of ceramic colors or stains for ceramic tile and sanitary wares. Zirconia and silica are fired together to form zircon in the presence of small amounts of other elements which are trapped in the zircon lattice to form colors such as tin–vanadium yellow, praseodymium–zircon yellow [68187-15-5], vanadium–zircon blue [12067-91-3], iron–zircon pink [68412-79-3], indium–vanadium orange (105–108).

Zirconium oxide increases the refractive index of some optical glasses, and is used for dispersion hardening of platinum and ruthenium. Very fine zirconium oxide has been used for polishing glass but ceria seems to be preferred.

Yttria or calcia-stabilized cubic zirconias are used extensively as the solid electrolyte in oxygen sensors in automotive and boiler exhausts, and oxygen-content probes for molten copper or iron in smelters. Stabilized zirconia has been proposed for high temperature fuel cells operating with natural gas, or converting ammonia to nitric acid (109). Stabilized zirconias are used for diverse applications such as ceramic tubing, extrusion dies, ball-milling media, high temperature heating elements (110), insulating fibers, thermal barrier coatings, ceramic diesel engines, and a nonlubricated ballbearing assembly for use in a space vehicle. Single-crystal cubic zirconias of high purity are being produced as a substitute for diamonds in jewelry (111) (see GEMSTONES).

Zirconate compounds exhibit several interesting properties. Lead zirconate–titanate [12626-81-2] compositions display piezoelectric properties which are utilized in the production of FM-coupled mode filters, resonators in microprocessor clocks, photoflash actuators, phonograph cartridges, gas

ignitors, audio tweeters and beepers, and ultrasonic transducers. Lanthanum-modified lead zirconate–titanate ceramics have been studied for photoferroelectric image storage (112) (see FERROELECTRICS). Alkaline-earth zirconate dielectrics are used in ceramic capacitors.

The uses of other zirconium compounds result primarily from the ability of zirconium to complex with carboxyl groups and to form an insoluble organic compound. Ammonium zirconium carbonate enhances the fungicidal action of copper salts on cotton cloth (113) and zirconium acetate has been used as an algicide. Both ammonium zirconium carbonate and zirconium acetate [14311-93-4] have been used in the waterproofing of fabrics (114). Freshly prepared zirconium sulfate solutions are preferred over chromium solutions in the tanning of leather to prepare white leathers, although other mixed sulfate solutions containing zirconium, chromium, and aluminum have been proposed. In the retanning of chrome-treated leather (qv), an intermediate treatment with ammonium zirconium carbonate improves the results obtained during the second chrome tanning (115). Zirconium carbonate produces durable acrylic-emulsion floor polishes which are easily stripped by treatment with household ammonia (see POLISHES). Both zirconium sulfate [14644-61-2] and potassium hexafluorozirconate have been used for flame-resistant treatment of wool fabric. Very dilute fluoride solutions containing zirconium and gluconic acids are used for the preparation of clean aluminum surfaces to be subsequently painted or overcoated (116). Similarly, an aqueous solution of zirconium nitrate [13746-89-9] and poly(vinyl alcohol) is applied as a precoating on glass surfaces before the application of phosphor photobinder in the production of cathode-ray tubes.

Zirconium tetrafluoride [7783-64-4] is used in some fluoride-based glasses. These glasses are the first chemically and mechanically stable bulk glasses to have continuous high transparency from the near uv to the mid-ir (0.3–6 μm) (117–118). Zirconium oxide and tetrachloride have use as catalysts (119), and zirconium sulfate is used in preparing a nickel catalyst for the hydrogenation of vegetable oil. Zirconium 2-ethylhexanoate [22464-99-9] is used with cobalt driers to replace lead compounds as driers in oil-based and alkyd paints (see DRIERS AND METALLIC SOAPS).

Sodium hydrogen zirconium phosphate [34370-53-1] is an ion-exchange material used in portable kidney dialysis systems which regenerate and recirculate the dialysate solution. The solution picks up urea during the dialysis. The urea reacts with urease to form ammonia, which is absorbed by the sodium hydrogen zirconium phosphate.

Zirconium phosphate [13772-29-7] also absorbs cesium and other radioactive-decay daughter products, and has been proposed as part of permanent disposal systems for nuclear fuel waste processing.

Zirconium metal is marketed in three forms: zirconium-containing silicon–manganese, iron, ferrosilicon, or magnesium master alloys; commercially pure zirconium metal; and hafnium-free pure zirconium metal. The use of zircon for the production of zirconium metal of all three types is ca 5–8% of the total U.S. zircon consumption.

Silicon–manganese–zirconium, ferrozirconium, and ferrosilicon–zirconium (and some pure zirconium) are used in the steel industry for deoxidizing. Magnesium–zirconium is added to magnesium and aluminum alloys for grain refining and strengthening. Most magnesium alloys used at elevated temperatures contain zirconium.

Because of its flammability, zirconium is used in military ordnance including percussion-primer compositions, delay fuses, tracers, and pyrophoric shrapnel, in getters for vacuum tubes, inert-gas glove boxes, and sodium-filled hollow-shaft exhaust valves, and as shredded foil in flashbulbs for photography and excitation of lasers. Alloying applications of pure zirconium include zirconium–niobium superconductors, titanium aircraft alloys, and strengthening of copper alloys. Pure zirconium is being increasingly used as a corrosion-resistant metal in fabricating columns, pumps, pipe, valves, heat exchangers, and tanks for severe chemical environments, particularly sulfuric and hydrochloric acids, except those containing fluorides. In this capacity, it is also used in facilities producing urea, hydrogen peroxide, methyl methacrylate, or acetic acid.

Hafnium-free zirconium alloys containing tin or niobium are used for tubing to hold uranium oxide fuel pellets inside water-cooled nuclear reactors. Zirconium–niobium alloys are used for pressure tubes and structural components in Canadian, the former USSR, and Germany reactor designs.

Hafnium-free zirconium is particularly well-suited for these applications because of its ductility, excellent oxidation resistance in pure water at 300°C, low thermal neutron absorption, and low susceptibility to radiation. Nuclear fuel cladding and reactor core structural components are the principal uses for zirconium metal.

Compounds

The numerous intermetallic compounds of zirconium, from ZrAl₃ to ZrZn₁₃, are reviewed in References 120–123.

Hydrides. Zirconium hydride [7704-99-6] in powder form was produced by the reduction of zirconium oxide with calcium hydride in a bomb reactor. However, the workup was hazardous and many fires and explosions occurred when the calcium oxide was dissolved with hydrochloric acid to recover the hydride powder. With the ready availability of zirconium metal via the Kroll process, zirconium hydride can be obtained by exothermic absorption of hydrogen by pure zirconium, usually highly porous sponge. The heat of formation is 167.4 J/mol (40 kcal/mol) hydrogen absorbed.

Zirconium absorbs or desorbs hydrogen reversibly. Provided the metal surface is clean, equilibrium is attained quickly above 400°C, and very slowly below 250°C. The hydrogen solubility limit in alpha zirconium has been expressed (124) as solubility, ppm = 1.61 × 10⁶exp (– 8959/RT). The addition of hydrogen lowers the α → β transition from 876 to 550°C at 6 at. % hydrogen. Further hydrogen addition causes the formation of hydride phases, culminating in the body-centered tetragonal epsilon phase as the hydrogen content approaches the limiting concentration of ZrH₂ (125–126). Hydrogen is removed by heating and evacuating. To reduce the hydrogen content to 10 ppm, the specimen must be heated to 900°C at 1 Pa (ca 10^{–2} mm Hg) (see HYDRIDES).

Zirconium hydride is not a true compound of fixed stoichiometry but rather a series of crystalline phases through which zirconium metal transforms with changing hydrogen concentration and temperature. The γ-phase hydride exists below 250°C in the narrow composition range ZrH_{0.9} to ZrH_{1.1}, the δ-phase has a compositional range ZrH_{1.5} to ZrH_{1.7}, and the ε-phase has a compositional range ZrH_{1.8}–ZrH₂. Most commercial hydride powder contains δ- and ε-phase.

Above 40 wt % hydrogen content at room temperature, zirconium hydride is brittle, ie, has no tensile ductility, and it becomes more friable with increasing hydrogen content. This behavior and the reversibility of the hydride reaction are utilized in preparing zirconium alloy powders for powder metallurgy purposes by the hydride–dehydride process. The mechanical and physical properties of zirconium hydride, and their variation with hydrogen content of the hydride, are reviewed in Reference 127.

Intermetallic compounds of zirconium with iron, cobalt, and manganese absorb and desorb considerable amounts of hydrogen, up to ZrMn_{2.4}H₃ [68417-38-9] (128) and ZrV₂H_{5.3} [63440-37-9] (129). These and other zirconium intermetallic compounds are being extensively studied for possible hydrogen storage applications (130).

The metallic monohalides zirconium chloride [14989-34-5], ZrCl₄, and zirconium bromide [31483-18-8], ZrBr₄, reversibly absorb hydrogen up to a limiting composition of ZrXH (131). These hydrides are less stable than the binary hydride ZrH₂, and begin to disproportionate above 400°C to ZrH₂ and ZrX₄ in a hydrogen atmosphere (see also HYDRIDES).

Carbide. Zirconium carbide [12020-14-3], nominally ZrC, is a dark gray brittle solid. It is made typically by a carbothermic reduction of zirconium oxide in a induction-heated vacuum furnace. Alternative production methods, especially for deposition on a substrate, consist of vapor-phase reaction of a volatile zirconium halide, usually ZrCl₄, with a hydrocarbon in a hydrogen atmosphere at 900–1400°C.

Once initiated, zirconium and carbon powders react exothermically in a vacuum or inert atmosphere to form zirconium carbide. With the greater availability of relatively pure metal powders, this technique is coming into common use for the production of several refractory carbides. Zirconium carbide is not a fixed stoichiometric compound, but a defect compound with a single-phase composition ranging from ZrC₀ to ZrC_{0.98} at 2400°C.

Zirconium carbide is inert to most reagents but is dissolved by hydrofluoric acid solutions which also contain nitrate or peroxide ions, and by hot concentrated sulfuric acid. Zirconium carbide reacts exothermically with halogens above 250°C to form zirconium tetrahalides, and with oxidizers to zirconium dioxide in air above 700°C. Zirconium carbide forms solid solutions with other transition-metal carbides and most of the transition-metal

nitrides without the formation of compounds. When heated in a vacuum, zirconium carbide, $\text{ZrC}_{0.98}$, moves to the congruent composition $\text{ZrC}_{0.85}$, with a melting point of 3420°C.

As a hard, high melting carbide and possible constituent of UC-fueled reactors, zirconium carbide has been studied extensively. The preparation, behavior, and properties of zirconium and other carbides are reviewed in Reference 132, temperature-correlated engineering property data in Reference 133 (see also CARBIDES).

Nitride. Cubic zirconium nitride, ZrN , a brittle, yellow solid, is prepared by heating zirconium turnings or loose sponge to 1100–1500°C in a nitrogen or ammonia atmosphere. The reaction is slow because of the slow diffusion of nitrogen through the protective nitride layer, and nitrogen contents approaching stoichiometric require time or higher temperatures such as those of a plasma torch. Zirconium tetrachloride and nitrogen is a hydrogen atmosphere above 1000°C yield zirconium nitride films or powders; impurity effects cause the growth of nitride whiskers (134). Ammonia and zirconium tetrachloride forms adducts which on heating to 750°C give $\text{ZrN}_{1.3}$ [25658-42-8], which in turn yields ZrN at 1200°C (135). Unlike its analogues, TiN and HfN , which are being used as wear-resistant surface layers on cemented-carbide tool bits, no significant commercial applications for ZrN have been developed (see also NITRIDES).

Zirconium nitride is dissolved by concentrated hydrofluoric acid, dissolved slowly by hot concentrated sulfuric acid, and oxidizes to zirconium oxide above 700°C in air.

Borides. Zirconium forms two borides: zirconium diboride [12045-64-6], ZrB_2 , and zirconium dodecaboride [12046-91-2], ZrB_{12} . The diboride is synthesized from the elements, by vapor-phase coreduction of zirconium and boron halides, or by the carbothermic reduction of zirconium oxide and boron carbide; boric oxide is avoided because of its relatively high vapor pressure at the reaction temperature.



The diboride has a hexagonal structure and melts at 3245°C (136); it is considered to have the best oxidation resistance of all the refractory hard metals. The dodecaboride has a cubic structure.

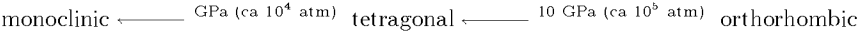
Phosphides. Zirconium forms several phosphides: ZrP_3 [39318-19-9], ZrP_2 [12037-80-8], and ZrP_0 [12066-61-4]; they are part of the Zr–P phase diagram (137). The solubility of phosphorus in zirconium metal is low, ca 50 ppm, and at higher concentrations it collects as separate globules at the metal grain boundaries. Analysis indicates that this material is Zr_3P .

Chalcogenides. The reactions of pure zirconium turnings with threefold quantities of elemental sulfur, selenium, or tellurium give ZrS_3 [12166-31-3], ZrSe_3 [12166-53-9], and ZrTe_3 [39294-10-5] (138). Zirconium disulfide [12039-15-5] is made from the elemental powders and by the action of carbon disulfide on zirconium oxide above 1200°C (139); some ZrOS [12164-95-3] is usually also obtained. The higher sulfides disproportionate at ca 700°C; synthesis reactions at 900–1000°C with S:Zr ratios between 0.2 and 2.3 produced crystals that were identified as Zr_3S_2 [12595-12-9], Zr_{21}S_8 [12595-19-6], Zr_2S [12334-07-5], $\text{ZrS}_{1.5}$, ZrS [12067-18-4], and $\text{ZrS}_{1.3}$ [37244-09-0] (140). Zirconium disulfide is a semiconductor with a cadmium–iodide-type layered structure consisting of stacked sandwiches each containing single sheets or metal cations between two sheets of anions.

Several compounds such as BaZrS_3 [12026-44-7], SrZrS_3 [12143-75-8], and CaZrS_3 [59087-48-8], have been made by reacting carbon disulfide with the corresponding zirconate at high temperature (141), whereas PbZrS_3 [12510-11-1] was produced from the elements zirconium and sulfur plus lead sulfide sealed in a platinum capsule which was then pressurized and heated (142). Lithium zirconium disulfide [55964-34-6], LiZrS_2 , was also synthesized. Zirconium disulfide forms organometallic intercalations with a series of low ionization (<6.2 eV)-sandwich compounds with parallel rings (143).

The dichalcogenides are hexagonal, the diamagnetic trichalcogenides are monoclinic. The compound sulfides BaZrS_3 [12026-44-7], SrZrS_3 [12143-75-8], and CaZrS_3 [59087-48-8] have an orthorhombic distorted perovskite structure, although BaZrS_3 prepared at 1000°C displayed a tetragonal perovskite structure which was attributed to a sulfur deficiency (141).

Oxides. Zirconium dioxide [1314-23-4], ZrO_2 , is the most important oxide of zirconium. It melts at $2710 \pm 35^\circ\text{C}$, and in the pure state exists in four solid phases: monoclinic, tetragonal, orthorhombic, and a cubic fluorite structure. The monoclinic phase is stable up to ca 1100°C, and transforms to tetragonal as the temperature increases to 1200°C. The proportion of tetragonal during the transformation is temperature dependent and not time-related (144). The transformation temperature exhibits a large hysteresis, and the tetragonal-to-monoclinic transition occurs between 1000 and 850°C; it is a martensitic transformation. The volume expansion resulting from the tetragonal-to-monoclinic transition can be restrained by high external pressure, increasing the stability of the tetragonal phase at lower temperatures (145). At higher pressures, an orthorhombic (cotunnite-type) structure has been identified (146) and it has been suggested that at 1000°C the zirconium oxide pressure-transformations are



The tetragonal-to-cubic reversible transformation occurs at 2370°C. The existence of cubic zirconia was proven by high temperature x-ray diffraction studies but little is known about the transformation. The cubic structure can be stabilized at lower emperatures by the introduction of vacancies in the anion lattice. The existence of cubic zirconia down to 1525°C has been shown in substoichiometric ZrO_{2-x} (147–149) and zirconium nitride-stabilized cubic zirconia (150). Vacancies are also generated by adding alkaline-earth oxides or rare-earth oxides to form stable solid solution cubic zirconia. Although these materials were initially considered stable at room temperature, more recent studies indicate that the kinetics of change are slow and that the lower temperature stability ranges from 280 to 1400°C, depending on the stabilizing oxide (151). Fully stabilized cubic magnesia–zirconia ceramics do not withstand thermal cycling, whereas CaO - and Y_2O_3 -stabilized oxides exhibit higher thermal stability.

Partially stabilized zirconias are formed by adding insufficient yttria, calcia, or magnesia to completely stabilize the zirconia, firing the mix, and then slowly cooling the sintered material in order to precipitate tetragonal zirconia within the grains of stabilized zirconia. The improved toughness of partially stabilized zirconias is the result of stress-induced martensitic transformation of the tetragonal particles to the monoclinic form in the stress field of a propagating crack. Their commercial application is based on their high strength and thermal shock resistance.

Zirconia prepared by the thermal decomposition of zirconium salts is often metastable tetragonal, or metastable cubic, and reverts to the stable monoclinic form upon heating to 800°C. These metastable forms apparently occur because of the presence of other ions during the hydrolysis of the zirconium; their stability has been ascribed both to crystallite size and surface energy (152–153) as well as strain energy and the formation of domains (154).

Zirconium oxide is stable to most reagents but dissolves slowly in hot concentrated sulfuric acid and in concentrated hydrofluoric acid. Carbon tetrachloride or phosgene above 300°C, or chlorine and carbon above 600°C give ZrCl_4 . Above 1400°C, zirconium oxide is reduced by carbon to zirconium carbide. At high temperatures, zirconium oxide reacts with many metal oxides, including BaO , CaO , MgO , SrO , Y_2O_3 , Re_2O_3 , and PbO , to form solid solution oxides and, with the exception of magnesium, zirconate compounds. Molten-salt synthesis procedures give uniform, stoichiometric, finely divided alkaline-earth zirconate powders for use as a dielectric (155).

Although lower oxides do not occur terrestrially, they can be produced at high temperature. Zirconium monoxide [12036-01-0], ZrO , has been observed in the sun's spectra, and in the spectra of zirconium oxide being evaporated from the surface of a tungsten filament (156). As oxygen is added to alpha zirconium, additional diffraction lines appear at Zr_5O [12059-93-7] and higher oxygen contents (157,158); Zr_5O occurs as an ordered super lattice in zirconium, representing the limit of oxygen solubility in zirconium. Other lower oxide compounds that are considered stable are Zr_8O [53801-45-9] and Zr_2O [12412-49-6] (159).

Silicates. Zirconium silicate, ZrSiO_4 , occurs naturally as zircon. Zircon is tetragonal, with the zirconium and silicon linked through oxygen atoms to form edge-sharing alternating SiO_4 tetrahedra and ZrO_8 triangular dodecahedra (160). Zircon is isomorphous with xenotime which occurs in solid solution in zircon crystals (161).

Upon heating to 900°C, zircon is transformed to a scheelite-type structure at 12 GPa (ca 120,000 atm), and a further transformation to the KAlF₄ structure at 17 GPa (ca 170,000 atm) has been predicted (162).

Zircon is synthesized by heating a mixture of zirconium oxide and silicon oxide to 1500°C for several hours (163). The corresponding hafnium silicate, hafnon, has been synthesized also. Zircon can be dissociated into the respective oxides by heating above 1540°C and rapidly quenching to prevent recombination. Commercially, this is done by passing closely sized zircon through a streaming arc plasma (38).

Zircon silicate is highly stable. Decomposition of zircon is accomplished only by very aggressive chemical attack, usually at high temperature.

Many other silicate minerals contain zirconium as a constituent (164,165). These minerals may be altered or metamict zircons such as cyrtolite or malacon.

Halides.

Quadrivalent. Zirconium tetrafluoride is prepared by fluorination of zirconium metal, but this is hampered by the low volatility of the tetrafluoride which coats the surface of the metal. An effective method is the halogen exchange between flowing hydrogen fluoride gas and zirconium tetrachloride at 300°C. Large volumes are produced by the addition of concentrated hydrofluoric acid to a concentrated nitric acid solution of zirconium; zirconium tetrafluoride monohydrate [14956-11-3] precipitates (69). The recovered crystals are dried and treated with hydrogen fluoride gas at 450°C in a fluid-bed reactor. The thermal dissociation of fluorozeirconates also yields zirconium tetrafluoride.

The physical properties of zirconium tetrafluoride are listed in Table 7.

Zirconium tetrafluoride dissolves in dilute acid without hydrolysis, and can be recovered as the monohydrate [14956-11-3] by crystallization from nitric acid solutions. If the solution is acidified with hydrofluoric acid, ZrF₄·3H₂O [14517-16-9] crystallizes at 10–30 wt % HF; HZrF₅·4H₂O [18129-16-9] crystallizes at 30–35 wt % HF, and at higher HF concentrations H₂ZrF₆·2H₂O [12021-95-3] can be recovered.

Potassium hexafluorozeirconate, K₂ZrF₆, can be crystallized from a zirconium oxide chloride [7699-43-6] solution by addition of excess hydrofluoric acid and a stoichiometric amount of potassium fluoride. Industrially, it is produced by the fusion of zircon flour with potassium fluorosilicate as described above (31). Sodium fluorozeirconates can be produced by equivalent processes, or by adding sodium fluoride to a warm acidified solution of potassium fluorozeirconate to precipitate the less soluble sodium salt. Several different sodium salts can be precipitated, depending on the sodium fluoride concentration: NaZrF₅·H₂O [20982-58-5] is recovered from sodium fluoride concentration less than 0.21%, Na₂ZrF₆ [16925-26-1] is recovered from concentrations between 0.21 and 0.4%, and Na₅Zr₂F₁₃ [12022-20-7] from concentrations between 0.47 and 1.18%. At higher sodium fluoride concentrations, Na₃ZrF₇ [17442-98-7] precipitates. All of these solutions must remain acidic to prevent the formation of oxide fluorides. The solubilities and equilibrium phases for ZrF–Na(K,Rb,Cs)F–H₂O systems are given in Reference 168. For comparison with the aqueous system, the molten NaF·ZrF₄ salt system had the following stable phases: Na₃Zr₄F₁₉ [12140-35-1], Na₇Zr₃F₃₁ [12140-37-3], Na₃Zr₂F₁₁ [12140-29-3], Na₂ZrF₆, Na₅Zr₂F₁₃, and Na₃ZrF₇ (169). (In addition, Na₄HfF₈ has been found in the hafnium system.)

Table 7. Physical Properties of Zirconium Tetrahalides

Property	ZrF ₄	ZrCl ₄	ZrBr ₄	ZrI ₄
color	white	white	white	orange-yellow
melting point, °C	932	437	450	499
sublimation temperature at 101.3 kPa (1 atm), °C	903	331	357	431
density, g cm ³	4.43	2.80		4.85
critical temperature, °C		503.5	532	686
critical pressure, MPa ^a		5.7	4.3	4.1
critical density, g cm ³		0.76	0.97	1.13
vapor pressure				
log P _{kPa} A B T ^b				
A	12.682	10.891	11.393	8.695
B	12430	5400	5945	6700
range t, °C	408–640	207–416	216–332	137–307
crystal structure	monoclinic (B)	monoclinic	cubic	cubic
space group, nm	1 2 a	P 2 c	P a3	P a3
a	0.957	0.6361	1.095	1.179 ^e
b	0.993	0.7407		
c	0.773	0.6256		
β°	99.47	109.30		
Zr–X bond distance, nm	0.210 avg	2.307 ^c		
		2.498 ^d		
		2.665 ^d		

^a To convert MPa to atm, divide by 0.101.
^b To convert kPa to mm Hg, multiply by 7.5. T = K.
^e Ref. 167.
^c Between terminal chlorine atoms.
^d Between bridging chlorine atoms.

Zirconium tetrachloride, ZrCl₄, is prepared by a variety of anhydrous chlorination procedures. The reaction of chlorine or hydrogen chloride with zirconium metal above 300°C, or phosgene or carbon tetrachloride on zirconium oxide above 450°C, or chlorine on an intimate mixture of zirconium oxide and carbon above 700°C are commonly used.

Zirconium tetrachloride is a tetrahedral monomer in the gas phase, but the solid is a polymer of ZrCl₄ octahedra arranged in zigzag chains in such a way that each zirconium has two pairs of bridging chlorine anions and two terminal or t-chlorine anions. The octahedra are distorted with unequal Zr–Cl bridge bonds of 0.2498 and 0.2655 nm. The physical properties of zirconium tetrachloride are given in Table 7.

Zirconium tetrachloride is instantly hydrolyzed in water to zirconium oxide dichloride octahydrate [13520-92-8]. Zirconium tetrachloride exchanges chlorine for oxo bonds in the reaction with hydroxylic ligands, forming alkoxides from alcohols (see ALKOXIDES, METAL). Zirconium tetrachloride combines with many Lewis bases such as dimethyl sulfoxide, phosphorus oxychloride and amines including ammonia, ethers, and ketones. The zirconium organometallic compounds are all derived from zirconium tetrachloride.

Zirconium tetrachloride forms additional compounds with phosphorus pentachloride: ZrCl₄·PCl₅ and ZrCl₄·2PCl₅. However, the alleged

2ZrCl₄·PCl₅ has been found to be a low boiling azeotrope of ZrCl₄ and ZrCl₄·PCl₅ (170).

Zirconium tetrachloride forms hexachlorozirconates with alkali-metal chlorides, eg, Li₂ZrCl₆ [18346-96-8], Na₂ZrCl₆ [18346-98-0], K₂ZrCl₆ [18346-99-1], Rb₂ZrCl₆ [19381-65-8], and Cs₂ZrCl₆, and with alkaline-earth metal chlorides: SrZrCl₆ [21210-13-9] and BaZrCl₆ [21210-12-8]. The vapor pressure of ZrCl₄ over these melts as a function of the respective alkali chlorides and of ZrCl₄ concentration were studied as potential electrolytes for the electrowinning of zirconium (72). The zirconium tetrachloride vapor pressure increased in the following sequence: Cs < Rb < K < Na < Li. The stability of a hexachlorohafnate is greater than that of a comparable hexachlorozirconate (171), and this has been proposed as a separation method (172).

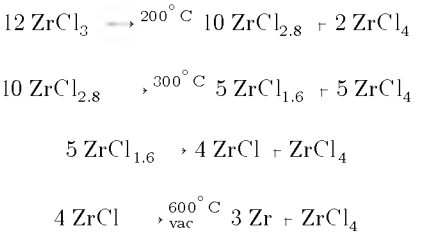
Zirconium tetrabromide [13777-25-8], ZrBr₄, is prepared directly from the elements or by the reaction of bromine on a mixture of zirconium oxide and carbon. It may also be made by halogen exchange between the tetrachloride and aluminum bromide. The physical properties are given in Table 7. The chemical behavior is similar to that of the tetrachloride.

Zirconium tetraiodide [13986-26-0], ZrI₄, is prepared directly from the elements, by the reaction of iodine on zirconium carbide, or by halogen exchange with aluminum triiodide. The reaction of iodine with zirconium oxide and carbon does not proceed. The physical properties are given in Table 7.

Zirconium tetraiodide is the least thermally stable zirconium tetrahalide. At 1400°C, it disproportionates to Zr metal and iodine vapor. This behavior is utilized in the van Arkel-de Boer process to refine zirconium. As with the tetrachloride and tetrabromide, the tetraiodide forms additional adducts with gaseous ammonia which, upon heating, decompose through several steps ending with zirconium nitride.

Lower Valent Halides. Zirconium trichloride [10241-03-9], ZrCl₃, tribromide [24621-18-9], ZrBr₃, and triiodide [13779-87-8], ZrI₃, are produced from zirconium metal and tetrahalide in sealed tantalum tubes at 500–700°C. The hafnium trihalides are made in the same manner. These reactions are slow, and may not reach equilibrium for days. Zirconium monohalides as the reductant increase the reaction rates. Another procedure uses quartz tubes with double bulbs; the halide is kept at the sublimation temperatures and the metal at 700°C. In a procedure employing comparatively low temperatures, the tetrahalides are dissolved in the corresponding liquid aluminum halide at the eutectic and reduced by metallic zirconium or aluminum between 220 and 300°C. The zirconium trihalides and ZrI_{3.40} [29950-62-7] were prepared in this manner (173). Zirconium trichloride, tribromide, and triiodide have been made from their respective tetrahalides by atomic hydrogen reduction (174).

The trichloride and tribromide disproportionate above 200°C at or below 101.3 kPa (1 atm) (175):



Other researchers have reported the trihalides to be nonstoichiometric with composition ranges of ZrCl_{2.94}–ZrCl_{3.03}, ZrBr_{2.87}–ZrBr_{3.23}, and ZrI_{2.83}–ZrI_{3.43} (176). The composition ranges of the lower iodides were reported to be ZrI_{1.9}–ZrI_{2.1}, and ZrI_{1.1}–ZrI_{1.3} (177).

Zirconium dichloride, ZrCl₂, has been made by a month-long reaction of ZrCl and ZrCl₃ at 650–750°C. The product has a three-layer sheet structure of close-packed Cl–Zr–Cl (178).

The first known cluster ion for group IVB elements is ZrCl₂,₆, produced similarly to ZrCl₂,: (Zr₆Cl₁₂)³⁺ (Cl[–])₃, with nine delocalized bonding electrons in the metal cage (178). ZrI₂ also contains the 12-electron cage ZrI₁₂, as does the isostructural ZrCl₁₂.

Attempts to prepare zirconium trifluoride, ZrF₃, by the zirconium reduction of zirconium tetrafluoride were unsuccessful, but it has been made by heating zirconium hydride to 750°C in a stream of hydrogen and hydrogen fluoride (179).

Monovalent Halides. Zirconium monochloride [14989-34-5], ZrCl, was discovered during electrorefining studies of zirconium in a SrCl₂–NaCl–ZrCl₄ melt intended to produce pure ductile hafnium-depleted zirconium from crude zirconium anodes (180–181). The monochloride is also called Zirklor. It is obtained as black flakes with a graphite slip-plane behavior and was proposed as a lubricant (182,183).

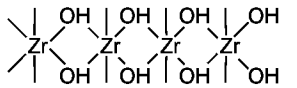
Zirconium monochloride and zirconium monobromide [31483-18-8] are prepared in better purity by equilibration of mixed lower halides with zirconium foil at 625°C (184–185) or by slowly heating zirconium tetrahalide with zirconium turnings at 400–800°C over a period of two weeks and holding at 800–850°C for a few additional days (186). Similar attempts to produce zirconium monoiodide [14728-76-8] were unsuccessful; it was, however, obtained from the reaction of hydrogen iodide with metallic zirconium above 2000 K (187).

Zirconium chloride and bromide have closely related but dissimilar structures. Both contain two metal layers enclosed between two nonmetal layers which both have hexagonal structure. In ZrCl, the four-layer sandwich repeats in layers stacked up according to abca bcab cabc, whereas the ZrBr stacking order is abca cabc bcab (188). Both are metallic conductors, but the difference in packing results in different mechanical properties; the bromide is much more brittle.

Zirconium monochloride is reported to disproportionate at 610°C in a high vacuum (184) and to exert a ZrCl₂ pressure of 101.3 kPa (1 atm) at 900°C and 293.7 kPa (2.9 atm) at 975°C, and to melt at 1100°C in an autogenous ZrCl₄ pressure of 1.5–2 MPa (15–20 atm) (185). The disproportionation proceeds at a lower temperature in the presence of copper or silver (182).

Zirconium monochloride reacts with sodium ethoxide to form additional adducts which hydrolyze in water. The monochloride does not react with benzene in a Friedel-Crafts reaction, and does not enter into intercalation reactions similar to those of zirconium disulfide. Both monohalides add hydrogen reversibly up to a limiting composition of ZrXH (131).

Hydroxyl Compounds. The aqueous chemistry of zirconium is complex, and in the past its understanding was complicated by differing interpretations. In a study of zirconium oxide chloride and zirconium oxide bromide, the polymeric cation [Zr₄(OH)₈(H₂O)₁₆]⁸⁺ was identified (189); the earlier postulated moiety [Zr=O]²⁺ was discarded. In the tetramer, the zirconium atoms are connected by double hydroxyl bridges (shown without the coordinating water molecules):



The tetramer exists in two-molal zirconium chloride and nitrate solutions, but it polymerizes into cross-linked chains on hydrolysis (190–191); in strong acid solutions, the hydroxyl bridges can be replaced by other anions to form trimers (192) and monomers (192–193).

Hydrous Oxides and Hydroxides. Hydroxide addition to aqueous zirconium solutions precipitates a white gel formerly called a hydroxide, but now commonly considered hydrous zirconium oxide hydrate [12164-98-6], ZrO₂·nH₂O. However, the behavior of this material changes with time and temperature.

The freshly precipitated material with four hydroxyl groups titratable with potassium fluoride for each zirconium atom can be considered a true hydroxide (194,195). When the freshly precipitated hydroxide was kept in water for two to three days before titration, two hydroxyl groups were titrated quickly, and two slowly on mixing for 30 h. These were called α- and γ-hydroxides. An intermediate β-hydroxide with three hydroxyl groups per zirconium could be produced only by precipitation from solutions of zirconium oxide chloride or nitrate in methanol, with extremely high zirconium concentrations. Drying these hydroxides in air gave a powder, which no longer contained any titratable hydroxyl groups; the material was very sparingly soluble in acid.

The hydroxides as precipitated are amorphous, but if they are refluxed in a neutral or slightly acidic solution they convert to a mixture of cubic and monoclinic hydrous zirconia crystallites; on continued refluxing, only the monoclinic form persists (196). If the refluxing is conducted in an alkaline solution, metastable cubic zirconia is formed (197).

Oxide Chlorides. Zirconium oxide dichloride, $\text{ZrOCl}_2 \cdot 8\text{H}_2\text{O}$ [13520-92-8], commonly called zirconium oxychloride, is really a hydroxyl chloride, $[\text{Zr}_4(\text{OH})_8 \cdot 16\text{H}_2\text{O}]\text{Cl}_8 \cdot 12\text{H}_2\text{O}$ (189). Zirconium oxychloride is produced commercially by caustic fusion of zircon, followed by water washing to remove sodium silicate and to hydrolyze the sodium zirconate; the wet filter pulp is dissolved in hot hydrochloric acid, and $\text{ZrOCl}_2 \cdot 8\text{H}_2\text{O}$ is recovered from the solution by crystallization. An aqueous solution is also produced by the dissolution and hydrolysis of zirconium tetrachloride in water, or by the addition of hydrochloric acid to zirconium carbonate.

Tetragonal prisms of $\text{ZrOCl}_2 \cdot 8\text{H}_2\text{O}$ crystallize from hot hydrochloric acid solutions by cooling or by increasing the acidity; in >32% hydrochloric acid transparent hexagonal plates are formed probably as $\text{ZrOCl}_2 \cdot 2\text{HCl} \cdot 10\text{H}_2\text{O}$ (198).

Zirconium oxychloride is an important intermediate from which other zirconium chemicals are produced. It readily effloresces, and hydrates with 5–7 H_2O are common. The salt cannot be dried to the anhydrous form, and decomposes to hydrogen chloride and zirconium oxide.

Zirconium hydroxy oxychloride [18428-88-1], nominally $\text{ZrO}(\text{OH})\text{Cl}$, is produced by dissolving hydrous zirconia in hydrochloric acid in an equal molar proportion, and is available only in solution. Other oxychlorides with Cl:Zr ratios <2 are discussed in Reference 199.

Anhydrous zirconium oxide chloride, ZrOCl_2 [7699-43-6], has been prepared by the reaction of dichlorine oxide with a zirconium tetrachloride suspension in carbon tetrachloride starting at 30°C and slowly rising to room temperature. The white solid is extremely hygroscopic and decomposes to ZrCl_4 and ZrO_2 at 250°C (200).

Nitrates. Anhydrous zirconium tetranitrate [12372-57-5], $\text{Zr}(\text{NO}_3)_4$, is prepared from zirconium tetrachloride and nitrogen pentoxide (201). The hydrated compounds are obtained from aqueous nitric acid (165); $\text{ZrO}(\text{NO}_3)_2 \cdot 2\text{H}_2\text{O}$ [20213-65-4] is most commonly used; $\text{Zr}(\text{NO}_3)_4 \cdot 5\text{H}_2\text{O}$ [12372-57-5] can be produced from strong nitric acid.

Aqueous solutions of zirconium oxide dinitrate [13826-66-9] (zirconium oxynitrate) and zirconium oxychloride behave very similarly; these two compounds have been cocrystallized in solid solution (202) where $\text{ZrO}(\text{NO}_3)_2 \cdot 5\text{H}_2\text{O}$ was the stable hydrate.

Carbonates. Basic zirconium carbonate [37356-18-6] is produced in a two-step process in which zirconium is precipitated as a basic sulfate from an oxychloride solution. The carbonate is formed by an exchange reaction between a water slurry of basic zirconium sulfate and sodium carbonate or ammonium carbonate at 80°C (203). The particulate product is easily filtered. Freshly precipitated zirconium hydroxide, dispersed in water under carbon dioxide in a pressure vessel at ca 200–300 kPa (2–3 atm), absorbs carbon dioxide to form the basic zirconium carbonate (204). Washed free of other anions, it can be dissolved in organic acids such as lactic, acetic, citric, oxalic, and tartaric to form zirconium oxy salts of these acids.

Basic zirconium carbonate is nominally $2\text{ZrO}_2 \cdot \text{CO}_2 \cdot x\text{H}_2\text{O}$ but the $\text{ZrO}_2:\text{CO}_2$ ratio may range from 4:1 to 1:1 depending on the methods and techniques of preparation.

Basic zirconium carbonate reacts with sodium or ammonium carbonate solutions to give water-soluble double carbonates. The ammonium double carbonate is nominally $\text{NH}_4[\text{Zr}_2\text{O}(\text{OH})_3(\text{CO}_3)_3]$. These solutions are stable at room temperature, but upon heating they lose carbon dioxide and hydrous zirconia precipitates.

In a study of zirconium double carbonates (205), a family of carbonates with $\text{CO}_3:\text{Zr}$ ratios of 1, 1.5, 2, 2.5, 3, 3.5, and 4:1 were identified. None of these compounds combined with ammonia, confirming the absence of HCO_3^- .

Sulfates. Sulfate ions strongly complex zirconium, removing hydroxyl groups and forming anionic complexes. With increasing acidity, all hydroxyl groups are replaced; zirconium sulfate [7446-31-3], $\text{Zr}(\text{SO}_4)_2 \cdot 4\text{H}_2\text{O}$, with an orthorhombic structure (206), can be crystallized from a 45% sulfuric acid solution. Zirconium sulfate forms various hydrates, and 13 different crystalline $\text{Zr}(\text{SO}_4)_2 \cdot n\text{H}_2\text{O}$ [14644-61-2] systems are described in Reference 207.

The tetrahydrate loses three hydration waters above 100°C and becomes anhydrous at 380°C. Many complex sulfates are formed with alkali-metal sulfates.

Basic zirconium sulfates are formed by hydrolysis of zirconium sulfate, which is broken up into fragments that undergo further hydrolysis to yield a series of basic sulfates with the generic formula $\text{Zr}_n(\text{OH})_{2n+2}(\text{SO}_4)_{n-1}$ (190).

Basic zirconium sulfate is also considered to be strands of $[\text{Zr}(\text{OH})_2^b]_n^{2n+}$ joined by bridging sulfates (208). The resulting formula is $\text{Zr}(\text{OH})_2^b(\text{OH})^t(\text{SO}_4)_{0.5}^b \cdot n\text{H}_2\text{O}$, where b are bridging anions and t are terminal anions.

The most common basic sulfate is $5\text{ZrO}_2 \cdot 3\text{SO}_3 \cdot n\text{H}_2\text{O}$ [84583-91-5] which is precipitated in good yield when a zirconium oxychloride solution is heated with the stoichiometric amount in sulfate ion. It is used to prepare high purity oxides and ammonium zirconium carbonate.

The stability of zirconium sulfate solutions to spontaneous precipitation when heated to 70°C for 2 h was studied as a function of $\text{SO}_3:\text{ZrO}_2$ ratio and metal concentrations (209). The zirconium solutions were considered unstable at metal concentrations below 0.64 M or at $\text{SO}_3:\text{ZrO}_2$ ratios <1.2.

Phosphates. Phosphate ions precipitate group IVB metals from strongly acid solutions. This ability is used in analytical procedures to separate zirconium from other elements, and to prepare zirconium phosphates. The precipitate is a gelatinous amorphous solid of variable composition. However, when refluxed in strong phosphoric acid, a crystalline, stoichiometrically constant compound forms of composition $\text{Zr}(\text{HPO}_4)_2 \cdot \text{H}_2\text{O}$, known as zirconium bis(monohydrogen phosphate) [13772-31-1] or α zirconium phosphate (210), or αZP . This compound is also obtained by gradual precipitation of the phosphate from a heated zirconium fluoride solution (211). On heating, the αZP undergoes three transformations, losing water and finally converting to cubic zirconium pyrophosphate [33712-62-8], ZrP_2O_7 , at 1000°C (212).

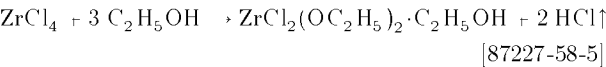
The main interest in zirconium phosphates relates to their ion-exchange properties. If amorphous zirconium phosphate is equilibrated with sodium hydroxide to pH 7, one hydrogen is displaced and $\text{ZrNaH}(\text{PO}_3)_2 \cdot 5\text{H}_2\text{O}$ [13933-56-7] is obtained. The spacing between the zirconium layers is increased from 0.76 to 1.18 nm, which allows this phosphate to exchange larger ions.

The gels precipitated as described above are not useful in ion-exchange systems because their fine size impedes fluid flow and allows particulate entrainment. Controlled larger-sized particles of zirconium phosphate are obtained by first producing the desired particle size zirconium hydrous oxide by sol–gel techniques or by controlled precipitation of zirconium basic sulfate. These active, very slightly soluble compounds are then slurried in phosphoric acid to produce zirconium bis(monohydrogen phosphate) and subsequently sodium zirconium hydrogen phosphate pentahydrate with the desired hydraulic characteristics (213,214).

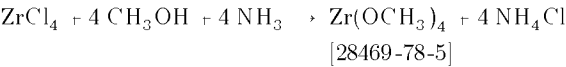
Although zirconium phosphate is insoluble in acids, it is easily hydrolyzed in excess caustic to give hydrous zirconium oxide. Zirconium phosphate forms soluble complexes with a large excess of zirconium oxide chloride, and therefore separation of phosphorus from zirconium oxide chloride solutions is difficult (215).

The properties and behavior of double phosphates such as $\text{Na}_{1+x}\text{Y}_x\text{Zr}_{2-x}(\text{PO}_4)_3$ (216), sodium–zirconium phosphate–silicates (217), and alkali-modified zirconium pyrophosphates (218) are under study as potential solid electrolytes in high temperature batteries.

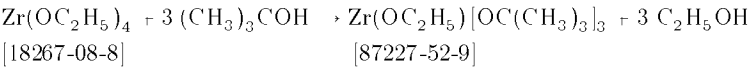
Alkoxides. Zirconium alkoxides are part of a family of alcohol-derived compounds (219). The binary zirconium compounds have the general formula $\text{ZrX}_4 \cdot n(\text{OR})_x$. They are easily hydrolyzed and must be prepared under anhydrous conditions. They are prepared by the reaction of zirconium tetrahalides and alcohols:



All four chlorines can be substituted if anhydrous ammonia is added to combine with the hydrogen chloride:



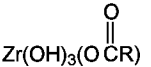
Alkoxides of other alcohols are formed by alcohol exchange. The general stability of the alcohols in exchange is primary > secondary > tertiary, although the reaction can be driven in the opposite direction by removal of the more volatile alcohol:



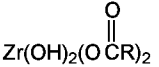
Alkoxides can be formed also by reaction of zirconium dialkylamines with alcohols, and alkoxides can be exchanged also be transesterification reactions. Double alkoxides of zirconium with alkali metals of the type $\text{MZr}_2(\text{OR})_6$ have been obtained by reaction of alkali metal alkoxides with zirconium alkoxides (220). Although these usually are monomeric derivatives, the reaction between zirconium tetra-*t*-butoxide [1071-76-7] and sodium *t*-butoxide was found (221) to form dimeric $[\text{NaZr}(\text{OC}(\text{CH}_3)_3)_5]_2$.

Zirconium alkoxides readily hydrolyze to hydrous zirconia. However, when limited amounts of water are added to zirconium alkoxides, they partially hydrolyze in a variety of reactions depending on the particular alkoxide (222). Zirconium tetraisopropoxide [2171-98-4] reacts with fatty acids to form carboxylates (223), and with glycols to form mono- and diglycolates (224).

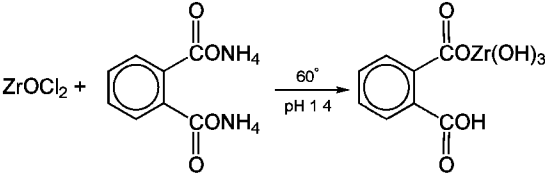
Carboxylates. Zirconium hydroxy carboxylates of the generic type



and

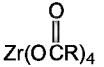


are precipitated by adding carboxylic salts to hydrochloric acid solutions of zirconium. With larger organic ligands, these compounds are water insoluble. They have been used for gravimetric determination, such as trihydroxyzirconium mandelate [87227-59-6] or for industrial purification precipitations, such as hydrogen trihydroxyzirconium phthalate [62313-97-7]:



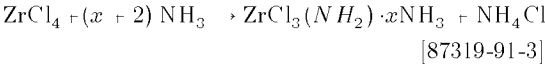
Zirconium oxalates exist as compounds, double compounds, and mixed oxalato complexes (165,195,225–226). When the carboxylate ligand is a longer alkyl chain, the materials often are called zirconium soaps.

Zirconium tetracarboxylates,



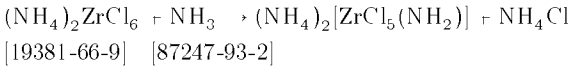
are prepared by reactions of zirconium tetraalkoxides or zirconium tetrachloride with anhydrous carboxylic acids (223).

Amides, Imides, Alkamides. When zirconium tetrachloride reacts with liquid ammonia, only one chloride is displaced to form a white precipitate, insoluble in liquid ammonia (227):

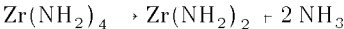
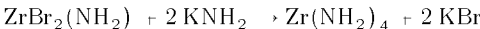


Upon heating in vacuum to 100°C, this material loses several ammonia moieties to give $\text{ZrCl}_3(\text{NH}_2) \cdot \text{NH}_3$.

The presence of NH_4Cl causes the formation of a complex which is soluble in excess liquid ammonia (228):

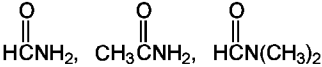


Additional halides can be removed from the above insoluble ammonolysis product by reaction with potassium imide to form zirconium imide [87227-54-1] $\text{Zr}(\text{NH}_2)_2$:

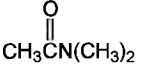


Zirconium dichlorobis(dimethylamide) [87227-57-4], $\text{ZrCl}_2[\text{N}(\text{CH}_3)_2]_2$ is made directly from methylamine and zirconium tetrachloride but all of the halogens can be substituted by treating zirconium tetrachloride with the appropriate lithium alkylamide (229). Zirconium arylamines are made from zirconium alkoxides which first are converted to aryloxides (230).

The zirconium–nitrogen bond is weaker than the zirconium–oxygen bond even under anhydrous conditions. When zirconium tetrachloride reacts with carbonyl-containing amides such as



or



all the bonding to zirconium is through carbonyl groups (231).

Organometallic Compounds. Certain zirconium organometallic compounds are highly reactive toward low molecular weight unsaturated

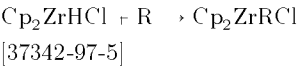
molecules. Some of these compounds are useful in various organic syntheses; others function effectively as catalysts for polymerization, hydrogenation, or isomerization.

Some simple zirconium organometallic compounds, such as tetramethylzirconium [6727-89-5], are known. In general, these compounds are very unstable. It appears that zirconium must be π-bonded to at least one moderately large ligand, such as a cyclopentadienyl group, for the compound to be stable. The abbreviation Cp is used here for the cyclopentadienyl group (C₅H₅), and Cp' for [C₅(CH₃)₅].

The metallocene complexes of M = Ti, Zr, and Hf are most stable when the two Cp groups are not parallel, in contrast to most other transition metal–Cp complexes. The most stable angle for the zirconium metallocenes is ca 40°, which partially accounts for the more interesting chemistry of these compounds compared to other transition metallocenes.

Cyclopentadienyl zirconium compounds are similar in structure and behavior to their titanium analogues. The increasing strength of the M–H and M–C bonds in the series M = Ti, Zr, and Hf makes the zirconium and hafnium compounds slightly more stable and thus somewhat less reactive than their titanium analogues. For example, ferrocenyl lithium reacts with Cp₂MCl₂ to give Cp₂M(Fc)₂. The colors of these compounds, green for Ti and red for Zr and Hf, are the results of electron transitions from a ferrocene orbital to a Cp₂M orbital (232), and is a measure of the relative bond energies. In general, the zirconium compounds are more complex and less reactive than the titanium analogues.

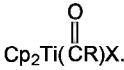
Hydrides. Zirconium hydrides react easily with unsaturated molecules. This process, termed hydrozirconation, replaces the hydrogen with the unsaturated group:



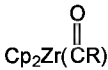
The oxidative addition reactions of Cp₂M(CO)₂ show some interesting differences between the Ti and Zr analogues. For Cp₂Ti(CO)₂, both (R = CH₃, C₆H₅)



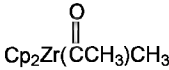
and RX(R = CH₃, C₂H₅, Pr⁺, Bu⁺) react to give acyls,



In contrast, Cp₂Zr(CO)₂ [59487-85-3] reacts with CH₃I, without CO insertion, to give Cp₂ZrCH₃I [63643-49-2]. Reaction with P(CH₃)₃ gives Cp₂Zr(CO) [P(CH₃)₃] [63637-45-6]; diphenylacetylene gives the metallocycle Cp₂Zr[C₄(C₆H₅)₄] [63637-34-3] (243,244).



R is formed by carbonylation of Cp₂ZrR₂ (R = CH₃ [12636-72-5], CH₂C₆H₅ [37206-41-0], and C₆H₅ [51177-89-0]) and is reversible for the alkyls, but not for aryls. Equilibrium and thermodynamic data were published for the alkyls. A crystal structure

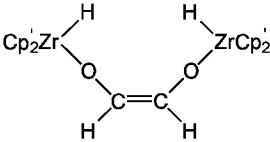


for

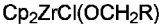


[60970-97-0] is used as support for interpreting the failure to eliminate the ketone as occurs in the Ti analogue, in terms of stabilization resulting from the side-bonded acyl group (245).

A Zr(IV) carbonyl, Cp₂ZrH₂(CO) [61396-35-8], was formed by exposing a solution of Cp₂ZrH₂ [61396-34-7] to CO at –80° C. The structure was inferred by observation of ¹³C–H coupling when ¹³CO was used. No ir spectra could be obtained. Warming Cp₂ZrH₂(CO) gives the dimer [61396-37-0] (246):



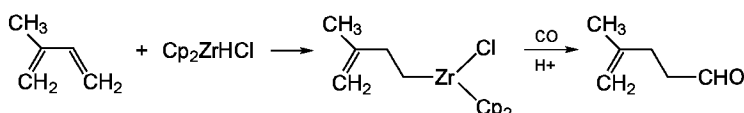
The CO stretching frequency of Cp'₂ZrH₂(CO) was observed by other workers at 2044 cm^{–1} (the Hf analogue was observed at 2036 cm^{–1}) (247). RCO₂H, R₂C=O, and RC≡N react to give Cp₂ZrXCl complexes where X = RCO₂[–], R₂CHO[–], and RCH₂N[–], respectively. Esters RCO₂R' give a mixture of



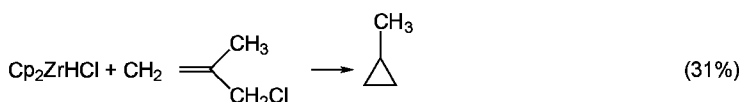
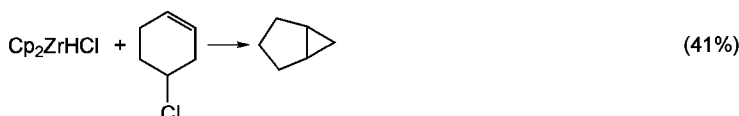
and Cp₂ZrCl(OR'); CO₂ is reduced to formaldehyde, with formation of (Cp₂ZrCl)₂O [12097-04-0]; cyclopentadiene produces Cp₂ZrCl [69005-93-2]; and ethylene gives Cp₂ZrCl(CH₂CH₃) [12109-84-1] (233).

The hydrides can also be used to form primary alcohols from either terminal or internal olefins. The olefin and hydride form an alkenyl zirconium, Cp₂ZrRCl, which is oxidized to the alcohol. Protonic oxidizing agents such as peroxides and peracids form the alcohol directly, but dry oxygen may also be used to form the alkoxide which can be hydrolyzed (234).

Zirconium hydrides undergo 1,2-addition with 1,3-dienes to give γ,δ-unsaturated complexes in 80–90% yield. Treatment of these complexes with CO at 20°C and 345 kPa (50 psi) followed by hydrolysis gives γ,δ-unsaturated aldehydes (235).



Cyclopropanes can be generated in high yield by treatment of halogen-containing alkenes with Cp_2ZrHCl ; eg,



Alkenes and alkanes are also produced (236). The procedures for the synthesis of Cp_2ZrH_2 [37342-98-6] and Cp_2ZrHCl have been published in detail (237).

Carbonyl Complexes. $\text{Cp}_2\text{M}(\text{CO})_2$ complexes ($\text{M} = \text{Ti}$, Zr [59487-85-3], and Hf) have been prepared by three different methods with varying yields: reduction of Cp_2MCl_2 (when $\text{M} = \text{Zr}$ [1291-32-3]) with $\text{Na}(\text{Hg})$ under CO at 101.3 kPa (1 atm) gives 80% with Ti , 9% with Zr , and 30% with Hf (238); reduction of Cp_2MCl_2 with Li under CO at 20 MPa (200 atm) gives 80% with Zr and 2% with Hf (239); and treatment of Cp_2MBH_4 with $(\text{C}_2\text{H}_5)_3\text{N}$ under CO at 101.3 kPa (1 atm) gives 80% with Ti and 15% with Zr (240). The nmr and ir spectra are similar for all three analogues, with the CO stretching frequencies decreasing in the order $\text{Ti} > \text{Zr} > \text{Hf}$ (238).

An x-ray study of the structure of $\text{Cp}_2\text{Hf}(\text{CO})_2$ revealed the expected tetrahedral disposition of ligands with OC-Hf-CO and (centroid Cp)- Hf -(centroid Cp) angles of 89.3° and 141° , respectively, and mean bond lengths for both bond types of 0.216 nm (241). The Zr analogue is isomorphous with bond lengths of 0.2187 nm and a OC-Zr-CO bond angle of 89.2° (242).

$\text{Cp}_2\text{Zr}(\text{CO})_2$ in hot toluene reacts with CO to give a cyclic trimer $(\text{Cp}_3\text{ZrO})_3$. The Zr_3O_3 ring in this unusual compound is nearly planar, with Zr-O-Zr and O-Zr-O bond angles of 142.5° and 97.5° , respectively. The mean Zr-O bond length (0.196 nm) seems to indicate a large degree of Zr-O double bonding (243).

Dinitrogen Complexes. The relative inertness of molecular nitrogen is well known, however, some $\text{Cp}'\text{-Zr}$ compounds coordinate dinitrogen and substantially increase its reactivity. The nitrogen molecule can be coordinated either in a terminal position or as a bridge in dimeric structures.

Bridging and terminal nitrogens have been compared in $[\text{Cp}_2\text{Zr}(\text{N}_2)]_2\text{N}_2$. The bridging N_2 has a longer N-N distance (0.118 vs 0.115 nm) and shorter Zr-N distances. In addition, the bridging N_2 stretching frequency is very low, 1578 cm^{-1} (248).

The increase in reactivity of coordinated N_2 has been assumed to be associated with increased bond length and decreased stretching frequency. A labeling study has shown that this is an oversimplification. In the protonolysis of $[\text{Cp}'_2\text{Zr}(\text{N}_2)]_2\text{N}_2$, the hydrazine produced comes equally from terminal and bridging N_2 . An intermediate, such as $\text{Cp}'_2\text{Zr}(\text{N}_2\text{H})_2$ [86165-22-2], was proposed where the bridging and terminal N_2 have become equivalent. Furthermore, careful carbonylation of the dimer produces $[\text{Cp}'_2\text{Zr}(\text{CO})]_2\text{N}$ which on protonolysis gives no reduced form of N_2 , indicating that both bridging and terminal N_2 are required for reduction (249).

A similar process has been patented covering $\text{Cp}'_2\text{ZrR}$ [86165-24-4], $\text{Cp}'_2\text{ZrR}(\text{N}_2)$ [86165-25-5], and $(\text{Cp}'_2\text{ZrR})_2\text{N}_2$ [86165-23-3], where $\text{R} = \text{CH}[\text{Si}(\text{CH}_3)_3]_2$. Protonolysis of the dinitrogen complexes gives hydrazine and ammonia (250).

Alkyl and Aryl Complexes. Cp_2MR_2 and Cp_2MRCl , [$\text{R} = \text{CH}(\text{C}_6\text{H}_5)_2$, $\text{CH}(\text{Si}(\text{CH}_3)_3)_2$] have been prepared and studied by nmr, and crystal structures for $\text{R} = \text{CHC}_6\text{H}_5$ have been reported. For Zr , the metal-carbon distance (0.2388 nm) is substantially longer than for $\text{R} = \text{CH}_3$, whereas no increase is observed with Hf (251).

The thermal decomposition of several Cp_2ZrR_2 compounds has been studied; RH is formed and the hydrogen derives from either a Cp or another R group and not from a solvent molecule (252).

Mixed-Metal Systems. Mixed-metal systems, where a zirconium alkyl is formed and the alkyl group transferred to another metal, are a new application of the hydrozirconation reaction. These systems offer the advantages of the easy formation of the Zr-alkyl as well as the versatility of alkyl-metal reagents. For example, Cp_2ZrRCl ($\text{R} = \text{alkyl}$ or alkenyl) reacts with AlCl_3 to give an Al-alkyl species which may then be acylated with



to give ketones,



in up to 98% yield. In contrast, direct acylation of the Zr compounds is difficult ($\text{R} = \text{alkyl}$) or impossible ($\text{R} = \text{alkenyl}$). The carbon configuration is retained during the transmetalation step which is faster for alkenyls, suggesting a bridged Zr-R-Al intermediate (253). Acyl-aluminum compounds,



which have potential use in organic synthesis as acyl anion equivalents, can be prepared by a similar reaction (254).

Alkenyl groups have also been transferred to Cu or Pd , leading to alkenyl dimers $\text{RCH}=\text{CH}-\text{CH}=\text{CHR}$ in high yield (255). Alkenyls undergo conjugate addition to enones in the presence of catalytic amounts of nickel acetylacetonate in high yield and selectivity (256). Nickel tetralithium catalyzes the reaction of Zr-alkenyls with aryl halides to give the cross-coupling products $\text{RCH}=\text{CHAr}$ (257). The reaction proceeds much more efficiently if the Ni catalyst is first reduced with diisobutyl aluminum hydride (258).

Alkenyl zirconium complexes derived from alkynes form C-C bonds when added to allylic palladium complexes. The stereochemistry differs from that found in reactions of corresponding carbanions with allyl- Pd in a way that suggests the Cp_2ZrRCl alkylates first at Pd , rather than by direct attack on the allyl group (259).

Catalysts. Several types of zirconium organometallic compounds are useful catalysts. In addition to the catalytic properties of the molecules, the fact that they can be bound to a relatively inert substrate increases their utility. A polymer-bound Cp-complex can be obtained through Cp–ring exchange. Treatment of Cp₂MCl₂ (M = Ti, Zr, Hf) with two equivalents of NaC₅D₅, followed by HCl, gives a mixture of (C₅H₅)₂–, (C₅H₅)(C₅D₅)–, and (C₅D₅)₂–MCl₂ (260). Although this indicates that ring exchanges in Cp₄M is fast, other studies have shown that the ratio of exchange products depends upon whether H- or D-substituted Cp₂MCl₂ is used as a reactant (261). Though Cp–ring exchange is not simply a matter of a Cp₄M intermediate, the products may nonetheless be used to bind Cp–Zr complexes to various substrates.

Polymer–Cp–MCl₃ complexes have been formed with the Cp-group covalently bound to a polystyrene bead. The metal complex is uniformly distributed throughout the bead, as shown by electron microprobe x-ray fluorescence. Olefin hydrogenation catalysts were then prepared by reduction with butyl lithium (262).

Ziegler polymerization catalysts may be prepared from Cp–Zr complexes and trialkylaluminum. The molecular weight of the polymers can be controlled over a wide range by varying the temperature. The activity of these catalysts is considerably increased by the addition of small amounts of water (263,264) (see OLEFIN POLYMERS).

Zirconium–allyl complexes also have catalytic properties. Tetraallylzirconium [12090-34-5] on a silica substrate catalyzes ethylene polymerization (265). Supported on silica, ZrR₄ (R = allyl or neopentyl) catalyzes olefin isomerization (266).

BIBLIOGRAPHY

"Zirconium and Zirconium Alloys" in *ECT* 1st ed., Vol. 15, pp. 282–290, by W. B. Blumenthal, National Lead Co.; "Zirconium Compounds" in *ECT* 1st ed., Vol. 15, pp. 290–312, by W. B. Blumenthal, National Lead Co.; "Zirconium and Zirconium Compounds" in *ECT* 2nd ed., Vol. 22, pp. 614–679, W. B. Blumenthal and J. D. Roach, National Lead Co.; "Zirconium and Zirconium Compounds" in *ECT* 3rd ed., Vol. 24, pp. 863–902, by R. H. Nielsen, J. H. Schlewitz, and H. Nielsen, Teledyne Wah Chang Albany.

1. N. E. Holden, *Pure Appl. Chem.* **52**, 2349 (1980).
2. B. Mason, *Principles of Geochemistry*, John Wiley & Sons, Inc., New York, 1952.
3. M. H. Klaproth, *Ann. Chim. Phys.* **6**(1), 1 (1789).
4. L. N. Vauquelin, *Ann. Chim. Phys.* **22**(1), 179 (1797).
5. J. J. Berzelius, *Ann. Chim. Phys.* **26**(2), 43 (1824).
6. H. D. Hess, *USBM R.I. 5856*, U.S. Bureau of Mines, Washington, D.C., 1962.
7. J. D. Grice and G. Perrault, *Can. Mineral.* **13**, 209 (1975).
8. H. G. Ansell, A. C. Roberts, and A. G. Plant, *Can. Mineral.* **18**, 201 (1980).
9. W. B. Pearson, *Handbook of Lattice Spacings and Structures of Metals and Alloys, International Series on Metal Physics and Physical Metallurgy*, Pergamon Press, New York, 1958, p. 130.
10. D. L. Douglass, *The Metallurgy of Zirconium*, International Atomic Energy Agency, Vienna, 1971, p. 4.
11. Ref. 10, p. 5.
12. O. Kubaschewski, ed., *Zirconium: Physico-Chemical Properties of its Compound and Alloys*, International Atomic Energy Agency, Vienna, 1976, p. 8.
13. Ref. 12, p. 9.
14. Ref. 12, p. 10.
15. Ref. 12, p. 11.
16. Ref. 10, p. 11.
17. Y. S. Touloukian, R. W. Powell, C. Y. Ho, and P. G. Klemens, *Thermophysical Properties of Matter*, Vol. 1, *Thermal Conductivity, Metallic Elements and Alloys*, IFI Plenum, New York, 1970.
18. B. Lustmann and F. Kerze, Jr., eds., *The Metallurgy of Zirconium*, McGraw-Hill, New York, 1955, p. 365.
19. Ref. 18, p. 370.
20. Ref. 10, p. 77.
21. Ref. 10, p. 78.
22. Ref. 18, p. 375.
23. J. H. McClain, *USBM IC 7686*, U.S. Bureau of Mines, Washington, D.C., 1954.
24. T. L. Yau, *Paper No. 102 presented at NACE Corrosion/81 April 1981*, Toronto, Canada.
25. H. L. Gilbert, C. Q. Morrison, A. Jones, and A. W. Henderson, *USBM RI 5091*, U.S. Bureau of Mines, Washington, D.C., 1954.
26. H. S. Choi, *Can. Min. Metall. Bull.* **67**, 65 (1965).
27. *Chem. Eng.* **61**, 124 (1954).
28. **U.S. Pat. 2,721,115** (Oct. 18, 1955) (to Zirconium Corp. of America); **2,721,117** (Oct. 18, 1955); **2,578,748** (Dec. 18, 1951) (to Sylvester & Co.); **3,832,441** (Aug. 27, 1974), R. A. Schoenlaub.
29. **U.S. Pat. 3,109,704** (1963), J. K. Olby.
30. G. von Hevesy, *Chem. Rev.* **2**, 1 (1925).
31. **U.S. Pat. 2,653,855** (1953), H. C. Kaweck.
32. N. P. Sajin and E. A. Pepelyaeva, in *Proceedings of the International Conference on Peaceful Uses of Atomic Energy*, Vol. 8, United Nations, New York, 1958, p. 559.
33. A. A. Manieh and D. R. Spink, *Can. Metall. Q.* **12**, 331 (1973).
34. A. A. Manieh, D. S. Scott, and D. R. Spink, *Can. J. Chem. Eng.* **52**, 507 (1972).
35. P. H. Wilks, *Chem. Eng. Prog.* **68**, (April 1974).
36. A. M. Evans and J. P. H. Williamson, *J. Mater. Sci.* **12**, 779 (1977).
37. R. H. Nielsen in D. E. Thomas and E. T. Hayes, eds., *Metallurgy of Hafnium*, U.S. Atomic Energy Commission, Washington, D.C., 1960, pp. 49–82.
38. I. V. Vinarov, *Russ. Chem. Rev.* **36**, 522 (1967).
39. D. Royston and P. G. Alfredson, *AAEC.TM 570*, Australian Atomic Energy Commission, Oct. 1970.
40. P. Pascal, *Nouveau traité de chimie minérale*, Tome IX, Masson, Paris, 1963.
41. O. A. Sinegribova and G. A. Yagodin, *At. Energy Rev.* **4**(1), 93 (1966).
42. W. Fischer and W. Chalybaeus, *Z. Anorg. Allg. Chem.* **255**, 79 (1947).
43. W. Fischer, W. Chalybaeus, and M. Zumbusch, *Z. Anorg. Allg. Chem.* **255**, 277 (1948).
44. J. H. McClain and S. M. Shelton in C. R. Tipton, Jr., ed., *Reactor Handbook*, Vol. 1, *Materials*, 2nd ed., Interscience Publishers, Inc., New York, 1960.
45. J. M. Googin in F. R. Bruce, J. M. Fletcher, and H. H. Hyman, eds., *Progress in Nuclear Energy, Series III: Process Chemistry*, Vol. 2, Pergamon Press, New York, 1958.
46. **U.S. Pat. 2,753,250** (July 3, 1956), H. A. Wilhelm, K. A. Walsh, and J. V. Kerrigan (to the United States of America as represented by the U.S. Atomic Energy Commission).
47. G. H. Beyer and H. C. Peterson, *AEC report ISC-182*, U.S. Atomic Energy Commission, Washington, D.C., Dec. 3, 1951.
48. R. P. Cox, H. C. Peterson, and G. H. Beyer, *Ind. Eng. Chem.* **50**(2), 141 (1958).
49. F. L. Moore, *National Academy of Sciences Report NAS-NS 3101*, Office of Technical Services, Department of Commerce, Washington, D.C., Dec.

- 1960.
50. E. Cerrai and C. Testa, *Energ. Nucl. (Milan)* **6**, 707 (1959).
51. Ref. 50, p. 768.
52. J. G. Moore, C. A. Blake, and J. M. Schmitt, *USAEC Report ORNL 2346*, U.S. Atomic Energy Commission, Washington, D.C., July 1957.
53. K. B. Brown, *USAEC Report ORNL-TM-107*, U.S. Atomic Energy Commission, Washington, D.C., Feb. 1962.
54. D. J. McDonald, *Sep. Sci. and Technol.* **16**, 1355 (1981).
55. M. Takahashi, H. Miyazaki, and Y. Katon, *paper presented at the 6th International Conference on Zirconium in the Nuclear Industry*, Vancouver, B.C., Canada, June 1982.
56. L. Moulin, P. Thouvenin, and P. Brun in Ref. 55.
57. **U.S. Pat. 4,021,531** (May 3, 1977), P. Besson, J. Guerin, P. Brun, and M. Bakes (to Ugine Aciers Pechiney Ugine Kuhlman, France).
58. **U.S. Pat. 2,852,446** (Sept. 16, 1958), M. L. Bromberg (to E. I. du Pont de Nemours & Co., Inc.).
59. *Chem. Eng.*, 106 (Jan. 26, 1981).
60. W. W. Minkler and E. F. Baroch in J. K. Tien and J. F. Elliott, eds., *Metallurgical Treatises*, Metallurgical Society of AIME, New York, 1981.
61. L. A. Nisel'son, T. D. Sokolova, and V. I. Stolyarov, *Dokl. Akad. Nauk SSSR* **168**, 385 (1966).
62. W. J. Kroll, A. W. Schlechten, and L. A. Yerkes, *Trans. Electrochem. Soc.* **89**, 263 (1976).
63. W. J. Kroll, A. W. Schlechten, W. R. Camody, L. A. Yerkes, H. P. Holmes, and H. L. Gilbert, *Trans. Electrochem. Soc.* **92**, 99 (1947).
64. W. J. Kroll, C. T. Anderson, H. P. Holmes, L. A. Yerkes, and H. L. Gilbert, *J. Electrochem. Soc.*, 94 (1948).
65. W. J. Kroll, W. W. Stephens, and H. P. Holmes, *Trans. AIME* **188**, 1445 (1950).
66. W. W. Stephens and J. H. McClain, *Chem. Eng.*, 116 (Mar. 1951).
67. S. M. Shelton, E. D. Dilling, and J. H. McClain, in Ref. 24 in *Proc. Int. Conf. on The Peaceful Uses of Atomic Energy*, Vol. 8, United Nations New York, 1957, pp. 505–550.
68. *Chem. Eng.*, 126 (Oct. 9, 1967).
69. W. J. S. Craigen, E. C. Joe, and G. M. Ritcey, *Can. Metall. Q.* **9**, 485 (1970).
70. *ZrHf Newsletter*, Amax Specialty Metals Corp., Akron, N.Y., Sept. 1973.
71. P. Pint and S. N. Flengas, *Trans. Instit. Min. Metallurgy* **87**, 29 (1978).
72. S. F. Flengas and P. Pint, *Can. Metall. Q.* **8**, 151 (1969).
73. Z. M. Shapiro in Ref. 18, Chapt. 5.
74. E. M. Sherwood and I. E. Campbell in Ref. 37, Chapt. 4.
75. V. S. Emelyanov, P. D. Bystrov, and A. I. Eustukhin, *J. Nucl. Energy* **4**, 253 (1957).
76. R. F. Rolston, *Iodide Metals and Metal Iodides*, The Electrochemistry Society, Inc., New York, 1961.
77. D. H. Baker, J. R. Nettle, and H. Knudsen, *USBM R.I. 5758*, U.S. Bureau of Mines, Washington, D.C., 1961.
78. J. C. Sehra, I. G. Sharma, P. L. Vijay, and A. Unnikrishnan, *Trans. Indian Inst. Met.* **30**, 14 (1977).
79. F. A. Schmidt, O. N. Carlson, and C. E. Swanson, Jr., *Metall. Trans.* **1**, 1371 (1970).
80. *Minerals Yearbook*, U.S. Bureau of Mines, Washington, D.C., 1972–1982.
81. F. D. Snell and L. S. Ettre, eds., *Encyclopedia of Industrial Chemical Analysis*, Vol. 14, Interscience Publishers, a division of John Wiley & Sons, Inc., New York, 1971, pp. 103–152.
82. R. B. Hahn in I. M. Kolthoff and P. J. Elving, eds., *Treatise on Analytical Chemistry*, Vol. 5, Interscience Publishers, a division of John Wiley & Sons, Inc., New York, 1961, Part II, pp. 61–138.
83. J. S. Fritz and M. Johnson, *Anal. Chem.* **27**, 1653 (1955).
84. P. R. Subbaraman and K. S. Rajan, *J. Sci. Ind. Res.* **13B**, 31 (1963).
85. C. L. Luke, *Anal. Chim. Acta* **37**, 284 (1967).
86. R. J. Jaworowski, R. P. Weveding, and D. J. Bracco, *Anal. Chim. Acta* **37**, 284 (1967).
87. *Methods for Emission Spectrochemical Analysis*, 7th ed., American Society for Testing and Materials, Philadelphia, Pa., 1982.
88. R. T. Van Santen, J. H. Schlewitz, and C. H. Toy, *Anal. Chim. Acta* **33**, 593 (1965).
89. E. B. Sandell, *Colorimetric Determination of Trace Metals*, 3rd ed., Interscience Publishers, New York, 1956, p. 976.
90. J. P. Yound, J. R. French, and J. C. White, *Anal. Chem.* **30**, 422 (1958).
91. Ref. 87, p. 766.
92. Ref. 87, p. 770.
93. J. H. Schlewitz and M. Shields, *At. Absorpt. Newsl.* **10**, 39 (1971).
94. S. V. Elinson and K. L. Petrov, *Zirconium, Chemical and Physical Methods of Analysis*, Main Administration on the Use of Atomic Energy at the Council of Ministers, Moscow, 1960; English translation, Office of Technical Services, Dept. of Commerce, Washington, D.C., 1962.
95. E. B. Read and H. M. Read, *The Chemical Analysis of Zirconium and Zircaloy Metals*, Nuclear Metals, Cambridge, Mass., 1957.
96. *Chemical Analysis of Zirconium and Zirconium-base Alloys*, ASTM E 146-68, American Society for Testing and Materials, Philadelphia, Pa., 1969.
97. W. B. Blumenthal, *J. Sci. Ind. Res.* **35**, 485, July 1976.
98. I. C. Smith and B. L. Carson, *Trace Metals in the Environment*, Vol. 3, *Zirconium*, Ann Arbor Science, Ann Arbor, Mich., 1978.
99. W. B. Shelley and H. J. Hurley in M. Sauter, ed., *Immunological Diseases*, 2nd ed., Little, Brown & Co., Boston, Mass., 1971.
100. W. W. Allison, *Zirconium, Zircaloy, and Hafnium Safe Practice Guide for Shipping, Storing, Handling, Processing, and Scrap Disposal*, WAPD-TM-17, Bettis Atomic Power Laboratory, Pittsburgh, Pa., 1960.
101. *National Fire Codes*, Vol. 13, NFPA No. 482M-1974, National Fire Protection Association, Boston, Mass., 1978.
102. J. Schemel, *ASTM Manual on Zirconium and Hafnium*, ASTM STP 639, American Society for Testing and Materials, Philadelphia, Pa., 1977.
103. *Bureau of Explosives Tariff No. BOE-6000-B*, Bureau of Explosives, Association of American Railroads, Washington, D.C.
104. R. A. Rowse and J. E. Patchett, *Sagamore Army Materials Research Conference Proceedings*, 1979, 1981, pp. 215–228.
105. C. A. Seabright and H. C. Draker, *Ceramic Bulletin* **40**, 1 (1961).
106. D. N. Crook, *J. Aust. Ceram. Soc.* **4**(2), 31 (1968).
107. C. L. Hackler, *Ceramic Bulletin* **54**, 283 (1975).
108. R. A. Eppler, *Ceramic Bulletin* **56**, 213 (1977).
109. R. D. Fare and C. G. Yayenas, *J. Electrochem. Soc.* **127**, 1478 (1980).
110. A. M. Anthony in A. H. Heuer and L. W. Hobbs, eds., *Science and Technology of Zirconia*, The American Ceramic Society, Columbus, Ohio, 1981, pp. 437–454.
111. K. Nassau, *Lapidary J.* **35**, 1210 (1981).
112. C. E. Land, *IEEE Trans. Electron Devices* **ED-26**, 1143 (Aug. 1979).
113. C. J. Conner, A. S. Cooper, L. W. Mazzeno, and W. A. Reaves, *Am. Dyest. Rep.*, 62 (April 1972).
114. W. B. Blumenthal, *Ind. Eng. Chem.* **42**, 640 (1950).
115. **U.S. Pat. 3,551,089** (Dec. 1970), W. O. Dawson, S. C. O'Connor, and F. S. Fellowes.
116. **U.S. Pat. 4,273,592** (June 1981), T. L. Kelly.
117. M. P. Brassington, T. Hailing, A. J. Miller, and G. A. Saunders, *Mater. Res. Bull.* **16**, 613 (1981).
118. R. N. Brown, B. Bendow, M. G. Drexhage, and C. T. Moynihan, *Appl. Opt.* **21**, 361 (1982).
119. N. W. Connon, *Eastman Org. Chem. Bull.*, Vol. 44 (1972).

120. M. Hansen, *Constitution of Binary Alloys*, McGraw-Hill, New York, 1958.
121. R. P. Elliott, *Constitution of Binary Alloys*, 1st Suppl., McGraw-Hill, New York, 1965.
122. F. A. Shunk, *Constitution of Binary Alloys*, 2nd Suppl., McGraw-Hill, New York, 1969.
123. O. Kubaschewski-von Goldbeck in Ref. 12, Chapt. 2.
124. Ref. 10, p. 158.
125. W. M. Mueller, J. P. Blackledge, and G. G. Libowitz, *Metal Hydrides*, Academic Press, New York, 1968, pp. 241–335.
126. E. Fromm and E. Gebhardt, *Gase und Kohlenstoff in Metallen*, Springer, Berlin, 1976.
127. R. L. Beck and W. M. Mueller in Ref. 125, pp. 286–321.
128. R. M. van Essen and K. H. J. Busehow, *Mater. Res. Bull.* **15**, 1149 (1980).
129. D. Shaltiel, I. Jacob, and D. Davidov, *J. Less Common Met.* **53**, 117 (1977).
130. D. G. Ivey, R. I. Chittim, K. J. Chittim, and D. O. Northwood, *J. Mater. Energy Syst.* **3**, 3 (Dec. 81).
131. A. W. Struss and J. P. Corbett, *Inorg. Chem.* **16**, 360 (1977).
132. E. K. Storms, *The Refractory Carbides*, Academic Press, New York, 1967.
133. *Engineering Property Data on Selected Ceramics, Vol. II Carbides*, Metal and Ceramics Information Center, Battelle Columbus Laboratory, Columbus, Ohio, 1979.
134. S. Motojima, E. Kani, Y. Takahashi, and K. Sugiyama, *J. Mater. Sci.* **14**, 1495 (1979).
135. Y. Okabe, Hojo, and Kato, *Yogyo Kyokai Shi* **85**, 173 (1977).
136. E. Rudy, *Compendium of Phase Diagram Material*, Air Force Materials Laboratory, Wright-Patterson Air Force Base, Ohio, 1969.
137. Ref. 12, p. 105.
138. L. Brattas and A. Kuekshus, *Acta Chem. Scand.* **26**, 3441 (1972).
139. H. Schafer, *Chemical Transport Reactions*, Academic Press, New York, 1964.
140. B. R. Conard and H. F. Franzen, *High Temp. Sci.* **3**, 49 (1971).
141. A. Clearfield, *Acta Crystallogr.* **16**, 135 (1963).
142. S. Yamoaka, *J. Am. Ceram. Soc.* **55**, 111 (1972).
143. W. B. Davies, M. L. H. Green, and A. J. Jackson, *J. Chem. Soc. Chem. Comm.* **19**, 781 (1976).
144. C. F. Grain and R. C. Garvie, *USBM R.I. 6619*, U.S. Bureau of Mines, Washington, D.C., 1965.
145. E. D. Whitney, *J. Electrochem. Soc.* **112**, 91 (1965).
146. L. G. Liu, *J. Phys. Chem. Solids* **41**, 331 (1980).
147. E. Gebhardt, H. Sechezzi, and W. Durshnabel, *J. Nucl. Mater.* **4**, 255 (1961).
148. R. Ruh and H. Garrett, *J. Am. Ceram. Soc.* **50**, 257 (1967).
149. R. J. Ackermann, S. P. Carg, and E. G. Rauh, *J. Am. Ceram. Soc.* **61**, 275 (1978).
150. N. Claussen, R. Wagner, L. J. Gauckler, and G. Petzow, *J. Am. Ceram. Soc.* **61**, 369 (1978).
151. U. S. Stubican and J. R. Hellman in Ref. 110.
152. R. G. Garvie in A. M. Alper, ed., *High Temperature Oxides*, Academic Press, New York, 1970, p. 127.
153. R. G. Garvie, *J. Phys. Chem.* **82**, 218 (1978).
154. H. T. Rijnten, *Zirconia*, Drukkerij Gebr. Janssen N. V. Nijmegen, 1971, pp. 67–70.
155. **U.S. Pats. 4,293,534; 4,293,535** (Oct. 6, 1981), R. H. Arendt (to General Electric).
156. E. G. Rauh and S. P. Garg, *J. Am. Ceram. Soc.* **63**, 239 (1980).
157. D. L. Douglass, *At. Energy Rev.* **1**, 75 (1963).
158. H. M. Chung and F. L. Yagee in Ref. 55.
159. Ref. 12, p. 104.
160. J. A. Speer in P. A. Ribbe, ed., *Orthosilicates*, Mineralogical Society of America, Washington, D.C., 1980.
161. P. A. Romans, L. Brown, and J. C. White, *Am. Mineralogist* **60**, 475 (1975).
162. R. M. Hazen and L. W. Finger, *Am. Mineral.* **64**, 196 (1979).
163. C. E. Curtis and H. G. Sowman, *J. Am. Ceram. Soc.* **36**, 190 (1953).
164. O. I. Lee, *Chem. Rev.* **5**, 22 (1928).
165. W. B. Blumenthal, *The Chemical Behavior of Zirconium*, D. Van Nostrand Co., Princeton, N.J., 1958.
166. H. D. Hess, *USBM R.I. 5856*, U.S. Bureau of Mines, Washington, D.C., 1962.
167. M. S. Ferrante and R. S. McClune, *BM RI 8418*, U.S. Department of the Interior, Washington, D.C., 1980.
168. I. V. Tananaev and L. S. Guzeeva, *Russ. J. Inorg. Chem.* **11**, 587 (1966).
169. C. J. Barton and co-workers, *J. Phys. Chem.* **62**, 665 (1958).
170. M. K. Chikanova, *J. Gen. Chem. USSR* **46**, 2088 (1976).
171. H. S. Ray, B. G. Bhat, G. S. Reddy, and A. K. Biswas, *Trans. Indian Inst. Met.* **32**, 177 (1978).
172. Can. Pat. 863,258 (Feb. 9, 1971), J. E. Dutrizag and S. N. Flengas.
173. E. M. Larsen, J. S. Wrazel, and L. G. Hoard, *Inorg. Chem.* **21**, 2619 (1982).
174. I. E. Newnham and J. A. Watts, *J. Am. Chem. Soc.* **82**, 2113 (1960).
175. R. A. J. Shelton, *J. Less Common Met.* **61**, 51 (1978).
176. R. L. Daake and J. D. Corbett, *Inorg. Chem.* **17**, 1192 (1978).
177. D. Cubicciotti, R. L. Jones, and B. C. Syrett in D. G. Franklin, ed., *Zirconium in the Nuclear Industry*, ASTM Special Technical Publication 754, American Society for Testing and Materials, Philadelphia, Pa., 1982.
178. J. D. Corbett, R. L. Daake, A. Cisar, and D. H. Guthrie in D. L. Hildenbrand and D. D. Cubicciotti, eds., *High Temperature Metal Halide Chemistry*, The Electrochemical Society, Princeton, N.J., 1978.
179. P. Ehrlich, F. Ploger, and E. Koch, *Z. Anorg. Allg. Chem.* **333**, 209 (1964).
180. **U.S. Pat. 2,941,931** (June 21, 1960), R. S. Dean.
181. Ref. 37, p. 69.
182. R. S. Dean, *Ind. Lab.* **45**, (April 1959).
183. F. L. Scott, H. Q. Smith, and I. Mockrin, *Technical Report No. 7, Office Naval Research Contract Nonr 2687(00)*, Pennsalt Chemicals Corp., Research and Development Dept., Wyndmoor, Pa., Sept. 10, 1959.
184. A. W. Struss and J. D. Corbett, *Inorg. Chem.* **9**, 1373 (1970).
185. R. L. Daake and J. D. Corbett, *Inorg. Chem.* **16**, 2029 (1977).
186. D. G. Adolfson and J. D. Corbett, *Inorg. Chem.* **15**, 1820 (1976).
187. P. D. Kleinschmidt, D. Cubicciotti, and D. L. Hildebrand, *J. Electrochem. Soc.* **125**, 1543 (1978).
188. J. F. Marchiando, B. N. Harmon, and S. H. Kiu, *Physica* **99B**, 259 (1980).
189. A. Clearfield and P. A. Vaughan, *Acta Crystallogr.* **9**, 555 (1956).
190. A. Clearfield, *Rev. Pure Appl. Chem.* **14**, 91 (1964).
191. J. R. Fryer, J. L. Hutchinson, and R. Paterson, *J. Colloid Interface Sci.* **34**, 238 (1970).
192. J. C. Mailen, D. E. Horner, S. E. Dorris, N. Pih, S. M. Robinson, and R. G. Yates, *Sep. Sci. Technol.* **15**, 959 (1980).
193. D. H. Devia and A. G. Sykes, *Inorg. Chem.* **20**, 910 (1981).

194. L. M. Zaitsev and G. S. Bochkarev, *Russ. J. Inorg. Chem.* **7**, 711 (1962).
195. L. M. Zaitsev, *Russ. J. Inorg. Chem.* **11**, 900 (1966).
196. A. Clearfield, *Inorg. Chem.* **3**, 146 (1964).
197. **U.S. Pat. 3,334,962** (Aug. 8, 1967), A. Clearfield (to National Lead Co.).
198. I. G. Kamaeva, L. A. Mel'Nik, and V. V. Serebrennikov, *Russ. J. Inorg. Chem.* **13**, 1026 (1968).
199. Ref. 165, pp. 132–135.
200. K. Dehnicke and J. Weidlein, *Angew. Chem. Int. Ed.* **5**, 1041 (1966).
201. B. O. Field and C. J. Hardy, *Proc. Chem. Soc.*, 76 (Feb. 1962).
202. F. M. Farhan and H. Nedjat, *J. Chem. Eng. Data* **21**, 441 (1976).
203. **U.S. Pat. 3,510,254** (May 5, 1970), R. N. Bell (to Stauffer Chemical Co.).
204. **U.S. Pat. 2,316,141** (April 6, 1943), E. Wainer (to the Titanium Alloy Manufacturing Co.).
205. L. A. Pospelova and L. M. Zaitsev, *Russ. J. Inorg. Chem.* **11**, 995 (1966).
206. J. Singer and D. T. Cromer, *Acta Crystallogr.* **12**, 719 (1959).
207. I. J. Bear and W. G. Mumme, *Rev. Pure Appl. Chem.* **21**, 189 (1971).
208. F. Famworth, S. L. Jones, and I. McAlpine in R. Thompson, ed., *Specialty Inorganic Chemicals*, The Royal Society of Chemistry, London, 1981, pp. 248–284.
209. M. A. Kolenkova and co-workers, *J. Appl. Chem. USSR* **50**, 156 (1977).
210. A. Clearfield and J. A. Stynes, *J. Inorg. Nucl. Chem.* **26**, 117 (1964).
211. G. Alberti and E. Torracca, *J. Inorg. Nucl. Chem.* **30**, 317 (1968).
212. A. Clearfield and S. P. Pack, *J. Inorg. Nucl. Chem.* **37**, 1983 (1975).
213. A. Clearfield, *Inorganic Ion Exchange Materials*, CRC Press, Boca Raton, Fla., 1982.
214. **U.S. Pat. 4,025,608** (May 24, 1977), D. S. Tawil, M. H. Clubley, and F. Famworth (to Magnesium Elektron Ltd.).
215. **U.S. Pat. 4,256,463** (Mar. 17, 1981), D. P. Carter (to Teledyne Industries, Inc.).
216. S. Fusitsu, M. Nagai, and T. Kanazawa, *Solid State Ionics* **3/4**, 233 (1981).
217. H. Y.-P. Hong, *Mater. Res. Bull.* **11**, 173 (1976).
218. R. Sacks, Y. Avigal, and E. Banks, *J. Electrochem. Soc.* **129**, 726 (1982).
219. D. C. Bradley, R. C. Mehrotra, and D. P. Gaur, *Metal Alkoxides*, Academic Press, New York, 1978.
220. W. G. Bartley and W. Wardlaw, *J. Chem. Soc.*, 421 (1958).
221. R. C. Mehrotra and M. M. Agrawal, *J. Chem. Soc.*, 1026 (1967).
222. Ref. 219, pp. 163–165.
223. R. N. Kapoor and R. C. Mehrotra, *J. Chem. Soc.*, 422 (1959).
224. U. B. Saxena, A. K. Rai, and R. C. Mehrotra, *Inorg. Chim. Acta* **7**, 681 (1973).
225. L. M. Zaitsev, *Russ. J. Inorg. Chem.* **9**, 1279 (1964).
226. L. M. Zaitsev, G. S. Bockharev, and N. V. Kozhenkova, *Russ. J. Inorg. Chem.* **10**, 590 (1965).
227. J. E. Drake and G. W. A. Fowles, *J. Less Common Met.* **2**, 401 (1960).
228. *Ibid.*, **3**, 149 (1961).
229. A. Allbutt and G. W. A. Fowles, *J. Inorg. Nucl. Chem.* **25**, 67 (1963).
230. **U.S. Pat. 3,626,010** (July 31, 1968), C. J. Sorrel (to Ethyl Corp.).
231. A. Clearfield and E. J. Malkiewich, *J. Inorg. Nucl. Chem.* **25**, 237 (1963).
232. G. A. Razuvaev, G. A. Domrachev, V. V. Sharutin, and O. N. Suvarova, *J. Organometal. Chem.* **141**, 313 (1977).
233. P. Etievant, G. Tainturier, and B. Gautheron, *C. R. Acad. Sci. Ser. C* **283**, 233 (1976).
234. T. F. Blackburn, J. A. Labinger, and J. Schwartz, *Tetrahedron Lett.*, 3041 (1975).
235. C. A. Bertelo and J. Schwartz, *J. Am. Chem. Soc.* **98**, 262 (1975).
236. W. Tam and M. F. Rettig, *J. Organometal. Chem.* **108**, (1976).
237. P. C. Wailes and H. Weigold, *Inorg. Synth.* **19**, 223 (1979).
238. J. L. Thomas and K. T. Brown, *J. Organometal. Chem.* **111**, 297 (1976).
239. B. Demerseman, G. Bouquet, and M. Bigorgne, *J. Organometal. Chem.* **107**, c19 (1976).
240. G. Fachinetti, G. Fochi, and C. Floriani, *J. Chem. Soc. Chem. Commun.*, 230 (1976).
241. D. J. Sikora, M. D. Rausch, R. D. Rogers, and J. L. Atwood, *J. Am. Chem. Soc.* **101**, 5079 (1979).
242. J. L. Atwood, R. D. Rogers, W. E. Hunter, C. Floriano, G. Fachinetti, and A. Chiesi-Villa, *A. Inorg. Chem.* **19**, 3812 (1980).
243. G. Fachinetti, C. Floriani, and H. Stoecki-Evans, *J. Chem. Soc. Dalton Trans.* 2297 (1977).
244. B. Demerseman, G. Bouquet, and M. Bigorgne, *J. Organometal. Chem.* **132**, 223 (1977).
245. G. Fachinetti, G. Fochi, and C. Floriani, *J. Chem. Soc. Dalton Trans.*, 1946 (1977).
246. J. M. Manriquez, D. R. McAlister, R. D. Sanner, and J. E. Bercaw, *J. Am. Chem. Soc. Chem. Commun.*, 230 (1976).
247. J. A. Marsella, C. J. Curtis, J. E. Bercaw, and K. G. Caulton, *J. Am. Chem. Soc.* **102**, 7244 (1980).
248. R. D. Sanner, J. M. Manriquez, R. E. Marsh, and J. E. Bercaw, *J. Am. Chem. Soc.* **98**, 8351 (1976).
249. J. M. Manriquez, R. D. Scanner, R. E. Marsh, and J. E. Bercaw, *J. Am. Chem. Soc.* **98**, 3042 (1976).
250. **Brit. Pat. 2,018,267** (1979), M. F. Lappert, M. J. S. Gynane, and J. Jeffrey.
251. J. L. Atwood, G. K. Barker, J. Holton, W. E. Hunter, M. F. Lappert, and R. Pearce, *J. Am. Chem. Soc.* **99**, 6645 (1977).
252. G. A. Razuvaev, V. N. Latyaeva, L. I. Vyshinskaya, G. A. Vasileva, V. I. Khruleva, and L. I. Smimova, *Dokl. Akad. NAUK SSSR* **231**, 114 (1976).
253. D. B. Carr and J. Schwartz, *J. Am. Chem. Soc.* **99**, 638 (1977).
254. D. B. Carr and J. Schwartz, *J. Organometal. Chem.* **139**, c21 (1977).
255. M. Yoshifuju, M. J. Loots, and J. Schwartz, *Tetrahedron Lett.*, c21 (1977).
256. M. J. Loots and J. Schwartz, *J. Am. Chem. Soc.* **99**, 8045 (1977).
257. E. Negishi and D. E. Van Horn, *J. Am. Chem. Soc.* **99**, 3168 (1977).
258. J. Schwartz, M. J. Loots, and J. Kosugi, *J. Am. Chem. Soc.* **102**, 1333 (1980).
259. J. S. Temple and J. Schwartz, *J. Am. Chem. Soc.* **102**, 7382 (1980).
260. J. G. Lee and C. H. Brubaker, *J. Organometal. Chem.* **135**, 115 (1977).
261. O. Khan, A. Domaond, and J. P. Letourmeux, *J. Organometal. Chem.* **132**, 149 (1977).
262. E. S. Chandrasekaran, R. H. Grubbs, and C. H. Brubaker, Jr., *J. Organometal. Chem.* **120**, 49 (1976).
263. A. Anderson, H.-G. Cordes, J. Herwig, W. Kaminsky, A. Merck, R. Mottweiler, J. Pein, H. Sinn, and H.-J. Vollmer, *Angew. Chem.* **88**, 649 (1976).
264. J. Sinn, W. Kanimshky, H.-J. Vollmer, and R. Wolot, *Angew. Chem. Int. Ed. Engl.* **19**, 390 (1980).
265. V. A. Zakharov, V. K. Dudchenko, E. A. Paukshtis, L. G. Karakchiev, and Y. I. Yermakov, *J. Mol. Catal.* **2**, 421 (1977).
266. J. Schwartz and M. D. Ward, *J. Mol. Catal.* **8**, 465 (1980).

General References

D. J. Carden, M. F. Lappert, and C. L. Ralston, *Chemistry of Organozirconium and Hafnium Compounds*, Halsted Press, a division of John Wiley & Sons, Inc.,

New York, 1986.

E. Negishi and T. Takahashi, *Aldrichimica Acta* **18**(2), 31–48 (1985). Excellent for organochemistry of zirconium and hafnium.

ZONE REFINING

Zone refining is one of a class of techniques known as fractional solidification, in which a separation is brought about by crystallization of a melt without solvent being added (see also CRYSTALLIZATION) (1–8). Solid–liquid phase equilibria are utilized, but the phenomena are much more complex than in separation processes utilizing vapor–liquid equilibria. In most of the fractional-solidification techniques described in the article on crystallization, small separate crystals are formed rapidly in a relatively isothermal melt. In zone refining, on the other hand, a massive solid is formed slowly and a sizable temperature gradient is imposed at the solid–liquid interface.

Zone refining was developed in the early 1950s at Bell Telephone Laboratories in response to the need for extremely pure germanium. It was an essential step for the development of the transistor and thus for the entire electronics revolution. Before that time, purification was carried out by normal freezing, also called progressive freezing. A melt was slowly frozen, causing more impurities to be concentrated in the last melt to freeze. In progressive freezing, however, a single solidification did not give sufficient purification. Simply remelting the entire ingot would redistribute the impurities, thereby eliminating the purification of the first solidification. the impure end could be removed, but handling and cutting introduce new impurities. The problem was solved by melting only a small part of the material and passing this molten zone down the ingot. Subsequent zone passes increased purification without requiring handling or cutting or permitting excessive back-mixing.

Zone refining can be applied to the purification of almost every type of substance that can be melted and solidified, eg, elements, organic compounds, and inorganic compounds. Because the solid–liquid phase equilibria are not favorable for all impurities, zone refining often is combined with other techniques to achieve ultrahigh purity.

The high cost of zone refining has thus far limited its application to laboratory reagents and valuable chemicals such as electronic materials. The cost arises primarily from the low processing rates, handling, and high energy consumption owing to the large temperature gradients needed.

Actually, zone refining is only one of a class of techniques known as zone melting in which a molten zone is passed down a solid rod. Zone melting is used routinely to collect impurities in high purity materials, eg, silicon, in preparation for chemical analysis. Growth of bulk single crystals is an important application that includes commercial float-zoning of silicon crystals for the semiconductor industry (7). Floating-zone melting is being considered for use aboard the U.S. Space Shuttle because the absence of gravity might permit larger crystals to be grown with greater homogeneity.

Addition of solvent to the zone allows crystals to be grown that either decompose before melting (incongruent melting) or have a very high vapor pressure at the melting point. This technique has been called both temperature-gradient zone melting and the traveling-heater method of crystal growth. It has also been used for fabrication of semiconductor devices (9).

Continuous zone-refining techniques have been developed, both theoretically and experimentally (1,4,10–11).

Theory

Early zone-refining theory attempted to correlate the concentration of impurities with location.

Progressive Freezing. Although a rod geometry is not required, directional solidification in the Bridgman configuration is easiest to visualize. It is widely used for single-crystal growth and preparation of dendritic and eutectic composite materials.

To derive the concentration profile for progressive freezing, a material balance is employed for solidification of a small fraction dg of melt, as shown in Figure 1. Integration from the beginning of solidification gives (1,4,8):

$$\int_0^g (1 - g)^{-1} dg = -\ln(1 - g) = \int_{w_o}^{w_i} (w_l - w_s)^{-1} dw_l$$

(1)

where g is mass fraction already solidified, w_o is the original mass fraction of impurity, w_i is the mass fraction of impurity in the solid at g , and w_l is the average impurity concentration in the liquid melt. It has been assumed that no mass transfer occurs in the solid, which is nearly always true because of the absence of convective mixing and the very low diffusion coefficients. Without solid-state diffusion, the solid, once it is formed, does not know the melt is there. This is analogous to batch evaporation with the removal of the vapor as it forms.

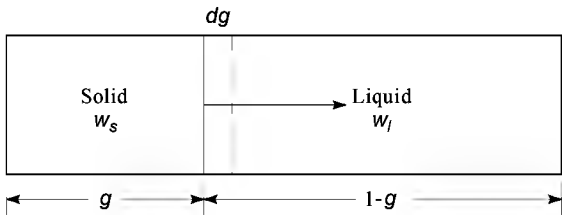


Fig. 1. Solidification of differential mass fraction dg of a melt. Mass fraction of impurity in melt is w_l and in solid freezing out is $w_s(g)$.

In order to integrate equation 1, it is necessary to have a relationship between w_l and w_s . If there is equilibrium between the bulk melt and the solid freezing out, phase-diagram data may be used. As shown in Figure 2, in the limit of very small impurity contents, w_s is often proportional to w_l , ie, $w_s = kw_l$ where k is the distribution or segregation coefficient. Equation 1 may then be integrated:

$$w_s - w_o = k(1 - g)^{k-1}$$

(2)

This relationship is plotted in Figure 3 for several values of k ; for $k < 1$ (the usual situation for impurities), $w_s \rightarrow \infty$ as $g \rightarrow 1$. Clearly, this assumption cannot be correct; w_s is no longer proportional to w_l at high concentrations.

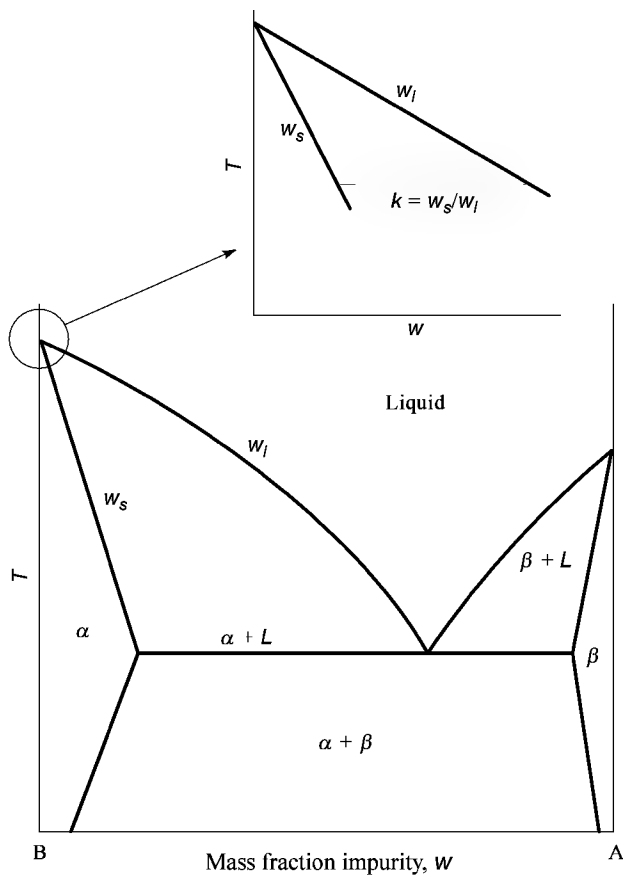


Fig. 2. Typical binary phase diagram for host and impurity, showing a constant distribution coefficient if impurity content is low. L = liquid composition after some solidification, α = B and small amount of A, β = A and small amount of B, w_l = liquidus, and w_s = solidus.

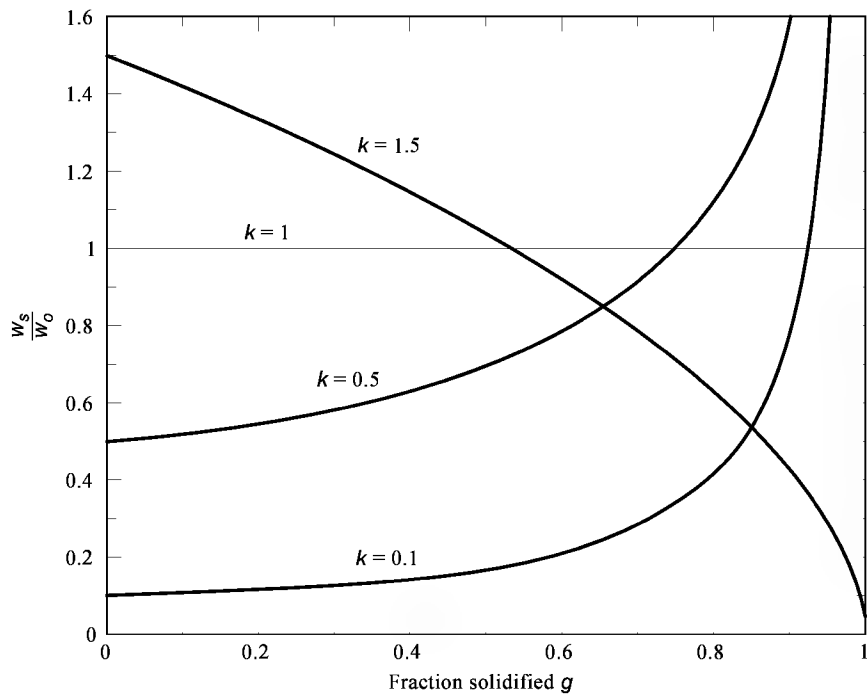


Fig. 3. Impurity concentration profiles resulting from progressive freezing with different values of distribution coefficient k (from eq. 2).

Zone Melting. A similar material balance may be made for a zone of mass m_z moving a short distance in such a manner that the mass dm of solid is frozen out and an identical mass melts into the zone (Fig. 4). For the first zone pass, it is assumed that the rod is initially at uniform composition w_o to obtain the following (1,4,8):

$$\int_0^m m_z^{-1} dm = m/m_z = \int_{w_o}^{w_s} (w_o - w_s)^{-1} dw_l \tag{3}$$

For constant $k = w_s/w_l$, this may be integrated to yield

$$w_s = w_o [1 - (1 - k) \exp(-km/m_z)] \tag{4}$$

Note that m/m_z is approximately the number of zone lengths down the rod.

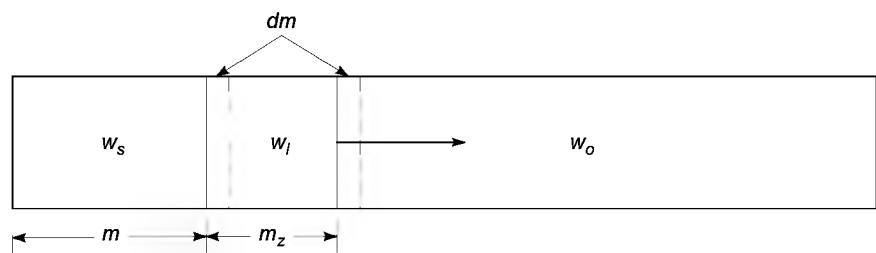


Fig. 4. Movement of molten zone by differential amount dm for initial zone pass; original solid has uniform impurity content w_o .

Computations are more difficult for subsequent zone passes, since the starting composition of the rod is no longer uniform. Nevertheless, a variety of numerical and analytical results has been obtained for infinite and for finite rods (1,4,12–16). A typical result is shown in Figure 5. Substantial purification can be attained even when k is not significantly different from 1.

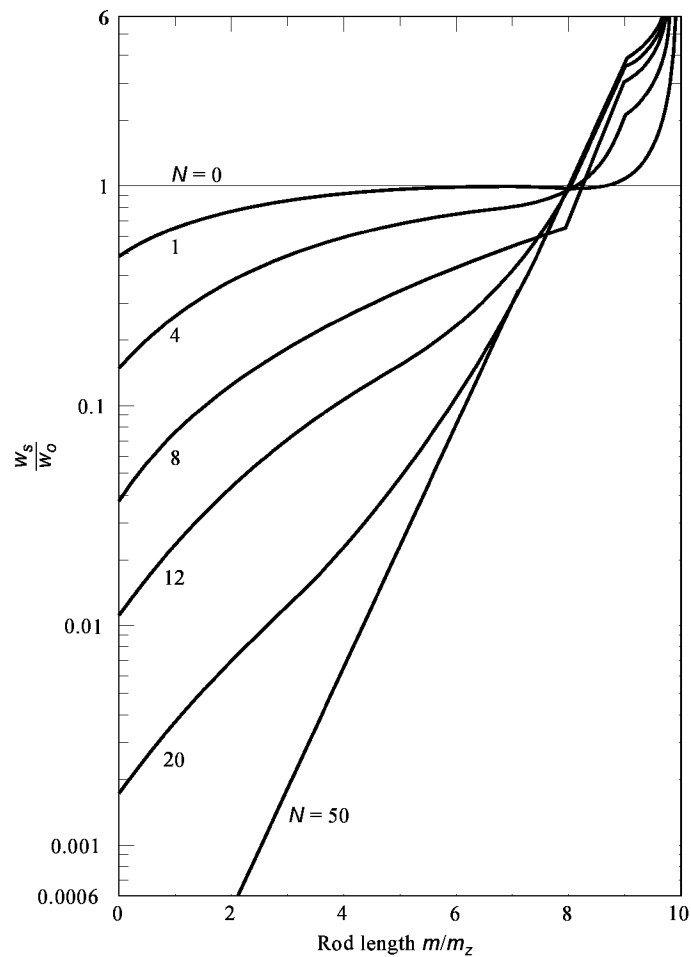


Fig. 5. Concentration profiles for different numbers N of zone passes for $k = 0.5$ and a rod containing 10 zone masses. Obtained by numerical methods of ref. 13 using Texas Instruments 99-4 Home Computer.

With subsequent zone passes, back-mixing becomes increasingly significant. When the number of zone passes becomes very large, the concentration profile reaches the ultimate distribution and subsequent passes have no influence. The number of zone passes required to approach the ultimate distribution is about twice the number of zone masses in the rod for small effective distribution coefficients. Thus, longer zones permit the attainment of the ultimate distribution sooner. On the other hand, the attainable separation increases as the zone size becomes smaller (permitting less back-mixing) (1,4).

Although most impurities have $k < 1$, there are exceptions, and it cannot be automatically assumed that the purest material is at the front end of the ingot; sometimes it is in the middle.

Theoretical treatments have also been given to other situations of interest. In zone leveling, the impurity level is homogenized either by moving the zone back and forth along a linear rod, or by movement around a ring (1). Evaporation and condensation of a volatile impurity (17–18), as well as decomposition of the material itself (13), have also been treated.

Nonideal Separations. In numerous instances, the ideal equations 2 and 4 have been verified experimentally. However, in other experiments different results were obtained, reflecting failure of one or more of the assumptions made in deriving equations 2 and 4. Likewise, much theoretical work is concerned with modified assumptions, including varying distribution coefficient k (19), eutectic-forming phase behavior (4,20–21), varying mass m_z of zone (22), and solid-state diffusion (23).

The assumption of equilibrium between solid w_s and bulk melt w_l is frequently violated because of lack of complete mixing in the melt. A steady-state fictitious stagnant-film treatment may be employed to arrive at an effective distribution coefficient,

$$k_{\text{eff}} = \frac{w_s}{w_l} = \frac{k}{k + (1 - k) \exp(-\delta V \rho_s / D \rho_l)} \tag{5}$$

where k is the equilibrium distribution coefficient from the phase diagram, δ is the film thickness (erroneously called the boundary-layer thickness), V the linear freezing rate, ρ_s the density of the solid, D the binary diffusion coefficient in the melt, and ρ_l the density of the liquid (1,4–5,24–26). The film thickness δ is related to the mass-transfer coefficient k_c and the Sherwood number N_{Sc} by

$$\delta = D \, k_c - L \, Sh$$

(6)

It has been found to be only weakly dependent on V (25). As shown in Figure 6, equation 5 predicts that $k_{\text{eff}} \rightarrow k$ as $V \rightarrow 0$, and $k_{\text{eff}} \rightarrow 1$ as $V \rightarrow \infty$. Thus, the separation obtained diminishes as the zone travel rate increases.

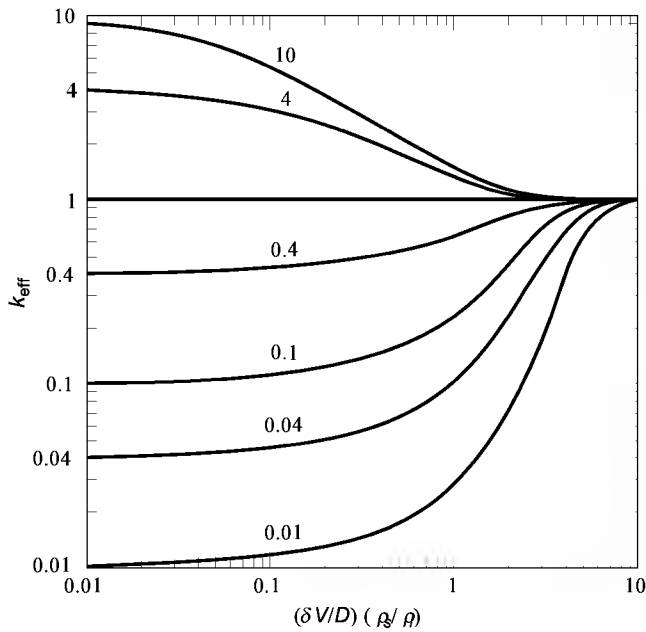


Fig. 6. Effective distribution coefficient k_{eff} vs interfacial distribution coefficient k and dimensionless zone-travel velocity; k_{eff} is to be used in place of k in expressions for concentration profiles, such as equations 2 and 4.

In practice, the freezing rate V is rarely constant. Fluctuating convective currents in the melt (27–32) lead to a fluctuating freezing rate, which causes k_{eff} to be somewhat nearer unity than predicted by equation 5. A fluctuating freezing rate also leads to impurity striations, ie, bands of varying impurity content parallel to the freezing interface. Irregular free convection in liquid metals has been damped successfully by means of a constant magnetic field, about 0.4 T (4 kG) for a melt of 1 cm dia (27,33–36).

The equilibrium distribution coefficient k varies with distance down the ingot because of compositional stress in the crystal (37–38), departure from linearity on the phase diagram at large impurity concentrations, and interactions with other impurities. The actual interfacial-distribution coefficient k_i can depart from equilibrium because of adsorption on the growing interface (5,39–40), differences in incorporation kinetics of impurity and host atoms (41–43), and interface-field effects (44). Deviation of k_i from k in such cases increases as the growth velocity V increases. For $k < 1$, k_i can even exceed 1 at moderate V . The well-known facet effect generally is thought to arise from such phenomena, and k_i is much greater on the faceted portion of the interface than on a nonfaceted part of the same interface (4–5).

As the zone-travel rate is increased, a breakdown of the freezing interface from smooth to dendritic is eventually observed. Dendrites are treelike structures that trap melt and dramatically reduce separation. The onset of interface breakdown is predicted by the elementary theory of constitutional supercooling (1,4–5). As shown in Figure 7a, $k < 1$ causes impurity to accumulate in the melt at the freezing interface, thereby lowering the freezing point. Thus, the actual temperature may be below the melting point for some distance into the melt, even though the two are identical at the freezing interface (see Fig. 7b). This is known as constitutional supercooling because it is brought about by a change in constitution or composition. The interface is unstable with respect to perturbation in shape, and cells or dendrites are expected.

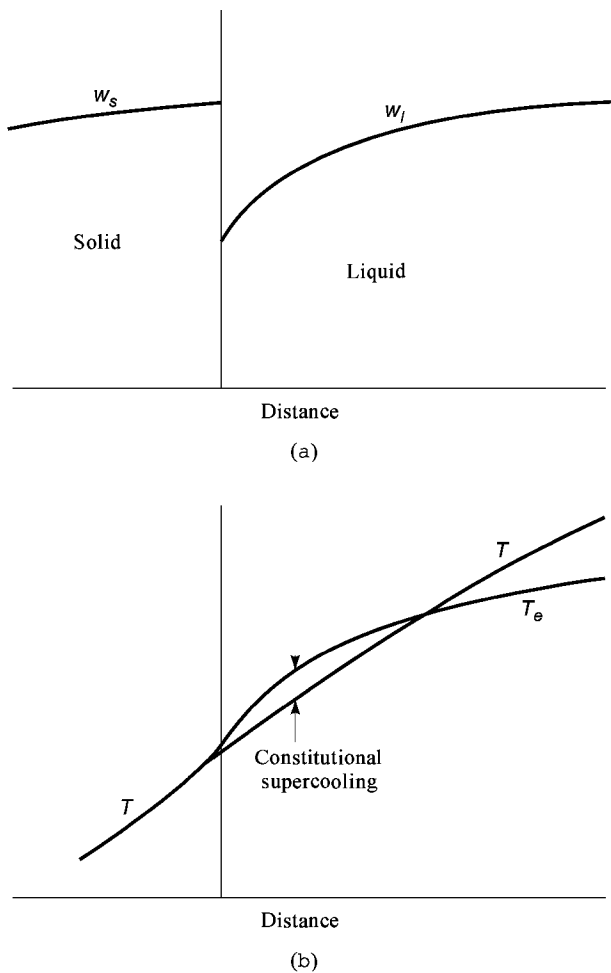


Fig. 7. Constitutional supercooling. (a) impurity concentration profile during solidification; (b) actual temperature T and equilibrium freezing temperature T_e during solidification.

At the onset of constitutional supercooling, the melting-point gradient exceeds the temperature gradient. Equating these gradients leads to the criterion for constitutional supercooling:

$$\begin{aligned} V/G > D\rho_l \left(\frac{\partial T_e}{\partial w_l} \right) \rho_s (w_s - w_l) - Dk\rho_l \left(\frac{\partial T_e}{\partial w_l} \right) \rho_s w_s (k - 1) \\ - D\rho_l [k + (1 - k) \exp(-\delta V \rho_s / \rho_l D)] \left(\frac{\partial T_e}{\partial w_l} \right) \rho_s (k - 1) w_l \end{aligned} \tag{7}$$

where G is the imposed interfacial-temperature gradient and $\partial T_e / \partial w_l$ is the slope of the liquidus on the phase diagram (change of freezing point with impurity concentration). Much more advanced and complicated treatments of morphological stability have been made (45–54). However, most experimental results are adequately predicted by the constitutional-supercooling theory outlined here (55–57).

From equation 7, it may be seen that the tendency toward constitutional supercooling increases as the freezing rate V increases, the temperate gradient G decreases, the impurity content w increases, the separation ($w_s - w_l$) between liquidus and solidus in the phase diagram increases, and the stirring decreases (δ increases). This explains why zone melting is limited to purification of materials with low impurity contents, and why substantial temperature gradients and low zone-travel rates are necessary.

Under the microscope, the onset of constitutional supercooling first manifests itself by the formation of a grooved interface, rather than dendrites. Dendritic growth occurs at higher freezing rates. In strongly faceted materials (high entropy of fusion), needles and plates may occur rather than branched dendrites. The separation is reduced whenever the interface is no longer smooth. Nevertheless, k_{eff} does not become precisely one even with severely dendritic growth, ie, some separation still occurs (58–61).

Insoluble particles may also be removed during zone refining. Most particles are pushed by a freezing solid–liquid interface, provided that the freezing rate is sufficiently slow. In the absence of stirring, a particle is pushed up to a critical freezing rate V_c , beyond which it is trapped (62–65). The critical freezing rate is roughly inversely proportional to the particle size for smooth particles and less dependent on particle size for rough particles. For 15- μm diameter smooth spheres, V_c is about 10 mm/h.

Stirring dramatically enhances separation of foreign particles by zone melting. If the particles are not allowed to rest on the freezing interface, they are pushed at much larger velocities than V_c in the absence of stirring, although there is no sharp demarcation velocity between pushing and engulfment. The probability of trapping increases as particle size and stirring decrease and the amount of particles increases (66–67).

A variety of other phenomena influence fractional solidification of organic compounds (68).

Optimization. Zone travel rate, sample geometry and orientation, zone length and spacing, stirring in the zone, and the number of zone passes can all be controlled. These variables can be optimized either with respect to purification or to the purification rate.

For maximum purification, the zone should move as slowly as possible, the sample should be many zone lengths long, the melt should be stirred, and more zone passes should be made than several times the number of zone lengths in the sample. Stirring and slow zone travel lower $\delta V/D$, which in turn increases $1 - k_{\text{eff}}$, as shown by equation 5 and Figure 6. In practice, it is sufficient for $\delta V/D \approx 0.1$.

Stirring is essential to maximize the rate of purification, and the travel rate V should be adjusted in such a way that $\delta V/D \approx 1$ (1,4). In addition, the narrow zones should be as close together as possible, a rod should be only about 10 zone lengths long, and there should be about as many zone passes as there are zone lengths in the rod. In the absence of tube rotation, heat transfer limits the minimum zone length and zone spacing to approximately the sample diameter, which is typically about 1 cm. The value $\delta V/D \approx 1$ derives from a consideration of equation 5, which is only valid in the absence of constitutional supercooling. In no case should the zone be moved so fast that dendritic growth occurs. Thus, whenever possible, the freezing interface should be examined with a microscope during zoning (69).

Dendritic growth can be avoided by keeping the zone travel rate low for the first few passes. As impurity is removed (for $k < 1$), higher zoning rates are possible. If dendritic growth is not a problem, $\delta V/D \approx 1$ at the beginning, followed by a gradual decrease of V for subsequent passes, in order for

$|1 - k_{\text{eff}}|$ to increase. Otherwise, the possible ultimate purification is limited, and the purification rate of later zone passes is low.

Since the separation occurs at the solid–liquid interface, the purification rate is maximized by having many zones close together. If this is not possible and only one zone can be passed at a time, the rate is maximized by having the first zone as long as possible and decreasing the length of subsequent zones.

Equipment and Techniques

Containers. The ideal container for zone melting should not contaminate the melt nor be damaged by the melt or subsequent contraction of the solid. For organic materials, borosilicate glasses are especially suitable, although metals and fluorocarbon and other polymers have also been successfully employed.

For many metals and semiconductors, fused silica (often erroneously called quartz) is ideal (5). Sometimes, the substance sticks to the glass and causes a break. Such breakage is prevented by coating the inside of the vessel with a film of carbon, conveniently deposited by passing acetone vapor through the glass tube at ca 700°C. However, at high temperatures such coating may result in contamination with carbon or silicon. Metals such as aluminum, with a high free energy of oxidation, react with silica glasses. In such cases, an oxide container of still higher free energy of formation is used, such as alumina or even sapphire. Zirconia and boron nitride have also been employed.

High melting inorganic salts and oxides are conveniently, albeit expensively, treated in noble-metal containers, platinum, rhodium, and iridium. Refractory containers, such as graphite, molybdenum, and tungsten in inert atmospheres, are also successfully employed for such salts. Low melting salts are refined in fused silica. Glassy carbon and pyrolytic graphite are good substitutes for ordinary graphite when carbon particles are a problem.

It is prudent to perform zone melting in a dry inert atmosphere. Oxygen causes most organic melts to oxidize slowly. Oxygen and moisture not only oxidize metals and semiconductors, but often enhance sticking to the container. Molten salts attack silica more rapidly in the presence of moisture. Oxygen and water are considered impurities in some inorganic compounds.

Moisture is removed from fused silica by heating, preferably in vacuum. Prior to this, silica ampuls are sometimes rinsed with a detergent solution, hydrofluoric acid, nitric acid, or methanol. After cleaning, drying, and loading, the ampul may be alternately evacuated and back-filled with inert gas. Sealing is easier if the inert-gas pressure is only slightly below atmospheric.

The container may either be in the form of a tube or a horizontal boat. A boat is advantageous for removal of the substance after zone melting, but may not be feasible if the substance is volatile. Most organic compounds, for example, are too volatile to be zone refined in an open boat. After zone melting, a silica glass ampul may be removed by etching with hydrofluoric acid, carefully cutting lengthwise with a diamond saw, or tapping with a ball-peen hammer to propagate cracks.

Because the solid and melt differ in density, repeated zone passes can transport material along the container. If the substance expands upon melting, it tends to travel toward the front, and vice versa. With a horizontal boat, this can be avoided by tilting. The tilt can be adjusted in such a way that whenever a zone reaches the end of the ingot some melt overflows, thereby removing some impurity and preventing back mixing upon subsequent zone passes. Precautions must be taken with a sealed tube to prevent rupture of the container caused by material transport. With organic compounds that expand upon melting, provisions must be made for this expansion when the zone forms at the front end of the ampul by leaving an empty space at the front end, separated from the substance by a fluorocarbon polymer plug. Each time a new zone is formed, the plug is forced to move to take up the expansion.

Drive Mechanisms. Either the heaters or the sample may be moved. The optimal zone travel rates are typically rather slow, ie, ca 1 cm/h. Thus, electric-motor drives require gearing systems, resulting in an undesirably jerky stick-slip movement. This is avoided with double-rod supports with linear ball bearings and with low backlash gears, where tension on the moving piece pushes in the direction of motion. To ensure a smooth drive, it is useful to take time-lapse films during zoning. Even with the most stringent precautions, both low level mechanical-drive fluctuations and temperature fluctuations persist, and a zone-travel rate >1 mm/h is generally not considered practical. If constitutional-supercooling problems require a rate below this, other purification methods are preferable.

Pneumatic and hydraulic drive systems have also been used, although not widely.

If many zones are present in the sample simultaneously, it may be advantageous to move the ampul (or heaters) slowly by one zone spacing, and then rapidly backward to catch the next zone. By such a reciprocating action, many zones can be moved continuously through the sample without a bank of heaters longer than the sample.

Heating and Cooling. Heat must be applied to form the molten zones, and this heat much be removed from the adjacent solid material (4,70). In principle, any heat source can be used, including direct flames. However, the most common method is to place electrical resistance heaters around the container. In air, nichrome wire is useful to ca 1000°C, Kanthal to ca 1300°C, and platinum-rhodium alloys to ca 1700°C. In an inert atmosphere or vacuum, molybdenum, tungsten, and graphite can be used to well over 2000°C.

Conductors can be heated to very high temperatures by induction heating wherein the sample is surrounded by a water-cooled coil carrying high current at high frequencies. A frequency of 450 kHz is typical for good conductors with a diameter >1–2 cm. Floating-zone melting requires frequencies of several MHz. Basically, the high frequency electromagnetic field induces a current within the sample itself, which is then heated by the resistance losses. The lower the frequency and the lower the electrical resistance, the deeper the field penetrates. Contact with the specimen is not required, and there are no limits to the temperatures attainable. If the substance is not a conductor, it is still possible to use induction heating. A conducting ring, eg, graphite, is place between the sample and the high frequency coil. This susceptor is heated by the field, and in turn transmits heat to the sample by radiation, conduction, and convection.

Molten zones are also formed by radiant heating (71). The light source may be focused carbon arcs, xenon lamps, sunlight, or lasers. Very high temperatures have been achieved with all of these. For example, sapphire has been float-zoned in this manner, at over 2000°C.

An electron beam can be used for floating-zone melting of a conducting rod. For example, a heated tungsten ring is placed around the zone in a vacuum and made strongly negative. Electrons are emitted from the tungsten and deliver the entire voltage-drop energy to the zone. A positive connection is made to the end of the rod to drain the current.

Electric heaters have also been directly immersed in the molten zone. Zone refining has been accomplished with a single helical heater rotating in an annular sample space (71).

It is important to control temperature and power input of the heaters. A fluctuating heat input leads to a fluctuating freezing rate, thereby reducing separation. It can also cause breakage of the container because $\rho_l \neq \rho_s$. A variable autotransformer (eg, Powerstat or Variac) fed by a constant voltage transformer (eg, Sola) affords a convenient and cheap source of energy. The voltage is increased until the zone is the desired size. An oversized constant-voltage transformer does not give constant voltage. It is not necessary to actually know the temperature anywhere in the system.

Heat is often removed by simply allowing it to escape by convection, radiation, and conduction. However, such uncontrolled escape can lead to very large temperature fluctuations. It is better to surround the entire container, heaters and all, with a controlled-temperature cooled chamber. Even then, buoyancy-driven free convection from the ampul can lead to small temperature fluctuations. Jets of air or cooling water applied directly onto the ampul adjacent to the heater have been employed. Both temperature and flow rate of the coolant should be controlled.

Floating-Zone Melting. No completely satisfactory container material exists for many high melting materials. In such cases, floating-zone melting may be employed (1,5,7). Heat is applied to a vertical rod, usually by induction, electron-beam, or radiant methods. In early applications, the roughly cylindrical molten zone was held by surface tension, and rods of only a few mm in diameter could be zoned thereby. The demand for large-diameter silicon for semiconductor devices led to a modification. Typically, an induction-heating pancake coil is used with a diameter smaller than that of crystal. As shown in Figure 8, the upper feed rod melts to a conical shape, with the melt running down through the so-called eye-of-the-needle induction heating coil. The melt collects in a puddle and freezes onto the growing crystal below. The electromagnetic field of the induction heating coil supports the molten zone. In this way, dislocation-free single silicon crystals of >10 cm dia and almost 1 m long are grown commercially.

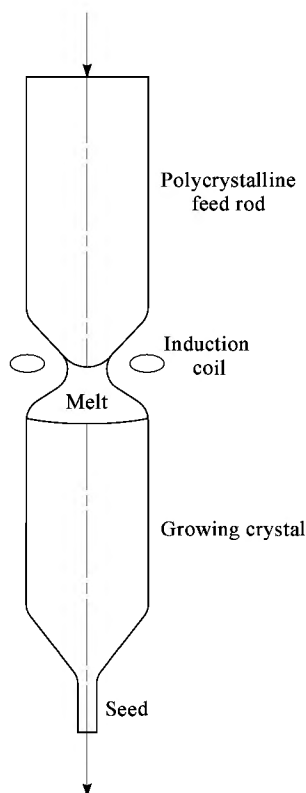


Fig. 8. Geometry for float-zoning large-diameter conducting materials.

The primary application for floating-zone melting is crystal growth rather than purification. Semiconductor-grade silicon is not purified by zone refining; silicon chlorides are distilled and then reduced with hydrogen.

Stirring. As noted earlier, stirring increases the optimal zone-travel rate by lowering δ and aids significantly in removal of foreign particles. Even in the absence of stirring, convection motion occurs in the melt (27). In a vertical sealed ampul, the driving force is buoyancy, ie, the interaction of gravity with density gradients owing to temperature and compositional variations. If the zone is shorter than the heater and the impurity content is low, the most vigorous convection is along the top interface. If that were the sole consideration, the top interface should be the freezing interface, ie, the zone should be moved downward. However, a gas bubble frequently forms at the top of the zone caused by release of the gas bubbles trapped in the solid during the casting of the feed ingot. Material evaporates across the gas space at the top of the zone, leading to a dramatically altered separation. Furthermore, the processes of evaporation, condensation, freezing, and dripping of condensate back into the melt are so complicated that the separation is difficult to predict and may be irreproducible (68). Therefore, the zone is frequently moved up so that the freezing interface is on the bottom.

The buoyancy-driven natural convection along the freezing interface in horizontal operation tends to be fairly vigorous. However, it also leads to spreading of the zone at the top owing to convection transport of heat upward.

With natural convection, δ is between 0.1 and several mm (4).

It is frequently very helpful to rotate the tube about its axis. In horizontal zone melting with a gas bubble along the top of the zone, rotation causes good stirring and enhances removal of foreign particles. Some insoluble particles are still trapped as long as the bubble contacts the freezing interface. Tilting the tube or making subsequent zone passes without a bubble can lead to virtually complete elimination of foreign particles. Although particles are very effectively removed if no gas bubble is present, δ is then about the same as with free convection (66–67).

With vertical zone melting and horizontal zone melting without a gas bubble, simple tube rotation at a constant moderate velocity does not significantly influence δ . In those cases, accelerated crucible rotation or spin up–spin down could be used (72–75). The tube is spun more rapidly than described above, but not at constant velocity. It may, for example, be spun rapidly, suddenly stopped, spun rapidly, etc, resulting in very vigorous stirring.

In recent years, it has been shown theoretically and experimentally that surface-tension gradients can give rise to very vigorous stirring (76–80). A free melt-vapor surface is required, as with floating-zone melting or horizontal processing in an open boat. Surface tension depends on temperature and composition, both of which vary along the melt surface. The surface moves from areas of low surface tension to areas of high surface tension, carrying the underlying melt along via viscosity. This is called Marangoni convection or thermocapillary convection. It occurs only if a surfactant does not immobilize the surface, which occurs surprisingly often. For example, most organic melts contain surfactant impurities which strongly inhibit Marangoni convection (81). Surface segregation also occurs in metal-alloy melts.

Induction heating is thought to cause vigorous convection because of the spatially varying average force field imposed along the melt surface.

Deliberate stirring can be imposed on conductors with a transverse rotating magnetic field or by passage of electric current axially with a transverse magnetic field. Conversely, a constant magnetic field with no current imposed greatly reduces natural convection.

Oscillators (82) and ultrasonic vibrations (83) have also been used to lower δ .

BIBLIOGRAPHY

"Zone Refining" in the *ECT* 2nd ed., Vol. 22, pp. 680–702, by Alan Lawley and D. Robert Hay, Drexel University; in *ECT* 3rd ed., Vol. 24, pp. 903–917, by W. R. Wilcox, Clarkson College of Technology.

1. W. G. Pfann, *Zone Melting*, 2nd ed., John Wiley & Sons, Inc., New York, 1966.
2. E. F. G. Herington, *Zone Melting of Organic Compounds*, John Wiley & Sons, Inc., New York, 1966.
3. H. Schildknecht, *Zonenschmelzen*, Verlag Chemie, Weinheim, Germany, 1964.
4. M. Zief and W. R. Wilcox, *Fractional Solidification*, Marcel Dekker, Inc., New York, 1967.
5. J. C. Brice, *The Growth of Crystals from Liquids*, North-Holland Publishing Co., Amsterdam, the Netherlands, 1973.
6. J. S. Shah in B. R. Pamplin, ed., *Crystal Growth*, Pergamon, Oxford, U.K., 1975, Chapt. 4.
7. W. Keller and A. Mshlbauer, *Floating-Zone Silicon*, Marcel Dekker, Inc., New York, 1981.
8. W. R. Wilcox, *CHEMI Module SMT-50*, American Institute of Chemical Engineers, in press.
9. M. Chang and R. Kennedy, *J. Electrochem. Soc.* **128**, 2193 (1981).
10. G. R. Atwood, *Chem. Eng. Prog. Symp. Ser.* **65**, 112 (1969).
11. A. M. Nazar and T. W. Clyne in T. N. Veziroglu, ed., *Proceedings Multi-Phase Flow and Heat Transfer Symposium*, Hemisphere, Washington, D.C., 1980.
12. N. W. Lord, *Trans. AIME* **197**, 1531 (1953).

13. W. R. Wilcox, *Sep. Sci. Technol.* **17**, 1117 (1982).

14. L. Burris, Jr., C. H. Stockman, and I. G. Dillon, *Trans. AIME* **203**, 1017 (1955).

15. E. Helfand and R. L. Kornegay, *J. Appl. Phys.* **37**, 2484 (1966).

16. I. Brown, *Brit. J. Appl. Phys.* **8**, 457 (1957).

17. J. R. Gould, *Trans. AIME* **221**, 1154 (1961).

18. Sh. I. Peizulaev, *Inorg. Mater.* **3**, 1329 (1967).

19. G. Matz, *Chem. Ing. Techn.* **4**, 381 (1964); Sh. I. Peizulaev, *Inorg. Mater.* **3**, 1329 (1967).

20. B. V. Ramarao, W. R. Wilcox, and W. E. Briggs, *J. Cryst. Growth* **59**, 557 (1982).

21. P. S. Ravishankar and W. R. Wilcox, *J. Cryst. Growth* **43**, 480 (1978).

22. H. M. Yeh and W. H. Yeh, *Sep. Sci. Technol.* **14**, 795 (1979).

23. D. Fischer, *J. Appl. Phys.* **44**, 1977 (1973).

24. J. A. Burton, R. C. Prim, and W. P. Slichter, *J. Chem. Phys.* **21**, 1987 (1953).

25. W. R. Wilcox, *Mater. Res. Bull.* **4**, 265 (1969).

26. W. R. Wilcox, *J. Appl. Phys.* **35**, 636 (1964).

27. J. R. Carruthers in W. R. Wilcox and R. A. Lefever, eds., *Preparation and Properties of Solid State Materials*, Vol. 3, Marcel Dekker, Inc., New York, 1977, Chapt. 1.

28. J. R. Carruthers, *J. Cryst. Growth* **32**, 13 (1976).

29. W. R. Wilcox and L. D. Fullmer, *J. Appl. Phys.* **36**, 2201 (1965).

30. M. A. Azouni, *J. Cryst. Growth* **47**, 109 (1979).

31. H. C. Gatos and A. F. Witt in A. Bishay, ed., *Recent Advances in Science and Technology of Materials*, Vol. 1, Plenum Publishing Corp., New York, 1973.

32. W. R. Wilcox, *Sep. Sci.* **2**, 411 (1967).

33. S. Sen, R. A. Lefever, and W. R. Wilcox, *J. Cryst. Growth* **43**, 526 (1978).

34. W. R. Wilcox and S. Sen, *Mater. Res. Bull.* **13**, 293 (1978).

35. H. P. Utech and M. C. Flemings, *J. Appl. Phys.* **37**, 2021 (1966).

36. A. F. Witt, C. J. Herman, and H. C. Gatos, *J. Mater. Sci.* **5**, 822 (1970).

37. W. R. Wilcox, *Mater. Res. Bull.* **2**, 121 (1967).

38. J. C. Brice, *J. Cryst. Growth* **28**, 249 (1975).

39. V. V. Voronkov and A. A. Chernov, *Sov. Phys.-Crystallogr.* **12**, 186 (1967).

40. D. E. Temkin, *Sov. Phys.-Crystallogr.* **17**, 405 (1972).

41. J. D. Weeks and G. H. Gilmer, *Adv. Chem. Phys.* **40**, 157 (1979).

42. H. Pfeiffer, *J. Cryst. Growth* **52**, 350 (1981).

43. J. C. Brice, *J. Cryst. Growth* **10**, 205 (1971).

44. W. A. Tiller and K.-S. Ahn, *J. Cryst. Growth* **49**, 483 (1980).

45. R. F. Sekerka in P. Hartman, ed., *Crystal Growth: An Introduction*, North-Holland Publishing Co., Amsterdam, the Netherlands, 1973, Chapt. 15.

46. D. J. Wollkind in W. R. Wilcox, ed., *Preparation and Properties of Solid State Materials*, Vol. 4, Marcel Dekker, Inc., New York, 1979, Chapt. 4.

47. S. R. Coriell and R. F. Sekerka, *J. Cryst. Growth* **34**, 157 (1976).

48. J. S. Langer, *Acta Met.* **25**, 1121 (1977).

49. T. Sato, K. Shibata, and G. Ohira, *J. Cryst. Growth* **40**, 69 (1977).

50. D. E. Temkin, *Sov. Phys. Crystallogr.* **22**, 529 (1977).

51. W. W. Mullins and R. F. Sekerka, *J. Appl. Phys.* **35**, 444 (1964).

52. A. L. Conlet, B. Billia, and I. Capella, *J. Cryst. Growth* **51**, 106 (1981).

53. S. R. Coriell, M. R. Cordes, W. J. Boettinger, and R. F. Sekerka, *J. Cryst. Growth* **49**, 13 (1981).

54. S. R. Coriell, D. T. J. Hurde, and R. F. Sekerka, *J. Cryst. Growth* **32**, 1 (1976).

55. D. E. Holmes and H. C. Gatos, *J. Appl. Phys.* **52**, 2971 (1981).

56. J. P. Dismukes and W. M. Yim, *J. Cryst. Growth* **22**, 287 (1974).

57. D. J. Morantz and K. K. Mathur, *J. Cryst. Growth* **16**, 147 (1972).

58. J. Verhoeven, *Met. Trans.* **2**, 2673 (1971).

59. D. D. Edie and D. J. Kirwan, *Ind. Eng. Chem. Fundam.* **12**, 100 (1973).

60. B. Ozum and D. J. Kirwan, *AIChE Symp. Ser.* **72**(153), 1 (1976).

61. N. Streat and F. Weinberg, *Met. Trans.* **7B**, 417 (1976).

62. A. A. Chernov, D. E. Temkin, and A. M. Mel'nikova, *Sov. Phys. Crystallogr.* **21**, 652 (1976).

63. S. N. Omenyi and A. W. Neumann, *J. Appl. Phys.* **47**, 3956 (1976).

64. R. R. Gilpin, *J. Colloid Int. Sci.* **74**, 44 (1980).

65. P. F. Aubourg, *Interaction of Second Phase Particles with Crystal Growing from Melt*, Ph.D. thesis, MIT, Cambridge, Mass., 1978.

66. J. E. Coon and W. R. Wilcox, *Sep. Sci. Technol.* **15**, 1401 (1980).

67. R. B. Fedich and W. R. Wilcox, *Sep. Sci. Technol.* **15**, 31 (1980).

68. W. R. Wilcox, *Sep. Sci.* **4**, 95 (1969).

69. C. E. Chang and W. R. Wilcox, *J. Cryst. Growth* **21**, 182 (1974).

70. N. Kobayashi, *J. Cryst. Growth* **43**, 417 (1978).

71. A. R. McGhie, P. J. Rennolds, and G. J. Sloan, *Anal. Chem.* **52**, 1738 (1980).

72. N. J. G. Bollen, M. J. Van Essen, and W. M. Smit, *Anal. Chim. Acta* **38**, 279 (1967).

73. G. J. Sloan, *Mol. Cryst.* **1**, 161 (1966).

74. K. D. Wolter, P. L. Carella, G. A. Moebus, and J. F. Johnson, *Sep. Sci. Technol.* **14**, 805 (1979).

75. I. N. Anikin and Kh. S. Bagdasarov, *Sov. Phys. Crystallogr.* **25**, 81 (1980).

76. P. A. Clark and W. R. Wilcox, *J. Cryst. Growth* **50**, 461 (1980).

77. C. E. Chang, W. R. Wilcox, and R. A. Lefever, *Mater. Res. Bull.* **14**, 527 (1979).

78. D. Schwabe and A. Scharmann, *J. Cryst. Growth* **46**, 125 (1979).

79. *Ibid.*, **52**, 435 (1981).

80. Ch.-H. Chun, *J. Cryst. Growth* **48**, 600 (1980).

81. W. R. Wilcox, R. S. Subramanian, J. M. Papazian, H. D. Smith, and D. M. Mattox, *AIChE J.* **17**, 1022 (1979).

82. B. K. Jindal, *J. Cryst. Growth* **16**, 280 (1972).

83. O. V. Abramov, I. I. Teumin, V. A. Filonenko, and G. I. Eskin, *Sov. Phys. Acoust.* **13**, 141 (1967).

William R. Wilcox
Clarkson College of Technology

ZYMURGY.

See BEER; FERMENTATION; WINE.